A DATA DRIVEN MINE-TO-MILL FRAMEWORK FOR MODERN MINES

by

Mustafa Erkayaoğlu

Copyright © Mustafa Erkayaoğlu 2015

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MINING AND GEOLOGICAL ENGINEERING

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015
As members of the Dissertation Committee, we certify that we have read the
dissertation prepared by Mustafa Erkayaoğlu, titled “A Data Driven Mine-to-Mill
Framework for Modern Mines” and recommend that it be accepted as fulfilling the
dissertation requirement for the Degree of Doctor of Philosophy.

Sean Dessureault

John Kemeny

Kwangmin Kim

Final approval and acceptance of this dissertation is contingent upon the candidate’s
submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and
recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Sean Dessureault

Dissertation Director: John Kemeny
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Mustafa Erkayaoğlu
ACKNOWLEDGEMENTS

I wish to acknowledge my deep sense of profound gratitude to my supervisor, Dr. Sean Dessureault for his illuminating and inspiring guidance and continuous encouragement throughout the course of the study.

I am deeply grateful to Dr. John Kemeny for his endless encouragement, supervision, suggestions, comments, and sincere guidance. Deep appreciations are extended to Dr. Young-Jun Son and Dr. Kwangmin Kim for being in my committee and for constructive criticism and valuable suggestions throughout the course of this study.

I am beholden to my father, Erdoğan Erkayaoğlu, my mother, Zeliha Erkayaoğlu, my sisters Arzu Yüzgeç and Alev Yazıcı, my brothers in law Ergin Yüzgeç and Hakan Yazıcı and my nephews Kaan Yazıcı, Azra Yüzgeç, and Aylin Yazıcı.
DEDICATION

To my family...
Table of Content

List of Figures ...................................................................................................................... 9
List of Tables ........................................................................................................................ 11
List of Acronyms ..................................................................................................................... 12
ABSTRACT ............................................................................................................................. 14

1. INTRODUCTION .............................................................................................................. 15
   1.1. Research objectives and goals ............................................................................... 15
   1.2. Research contributions ....................................................................................... 17
   1.3. Format of dissertation ...................................................................................... 18

2. LITERATURE REVIEW .................................................................................................. 19
   2.1. Mine-to-Mill approaches and practices worldwide ........................................... 19
   2.2. Blast fragmentation models in mine-to-mill optimization .................................. 21
   2.3. Data warehousing for mining related data ....................................................... 31
   2.4. Data mining for knowledge discovery in mining related data ....................... 34

3. METHODOLOGY OF DATA-DRIVEN MINE-TO-MILL OPTIMIZATION FRAMEWORK ............................................................. 37
   3.1. Data-driven Mine-to-Mill Optimization Framework ........................................... 37
   3.2. Mine-to-mill Optimization Based on Data Warehousing ................................. 41
      3.2.1. Integration of relational and process data through data warehousing .... 42
      3.2.2. Business process improvement for drilling and blasting practices ......... 44
   3.3. Mine-to-Mill Optimization through Data Mining ................................................ 48
      3.3.1. Data cleaning and modeling in Rattle ......................................................... 49
      3.3.2. Data mining models ............................................................................... 52
      3.3.3. Evaluation of data mining models ............................................................ 55

4. APPLICATION OF BUSINESS INTELLIGENCE AND DATA MINING .................................................. 58
   4.1. Information about the mine ............................................................................... 58
   4.2. Information about data sources available in the data warehouse .................. 60
4.3. Business Intelligence applications on mining related data

4.3.1. Geological conditions

4.3.2. Structure

4.3.3. Rock strength

4.3.4. Rock density

4.3.5. Blasthole design

4.3.6. Blasthole diameter

4.3.7. Burden

4.3.8. Spacing

4.3.9. Bench height

4.3.10. Sub-drilling

4.3.11. Stemming

4.3.12. Water depth

4.3.13. Material type

4.3.14. Mill type

4.4. Preparation of mining related data for data mining applications

4.5. Derivation of innovative parameters for data mining

4.5.1. Geospatial data manipulation of drill navigation system data

4.6. Deployment of the data mining model to the data warehouse

5. RESULTS OF DATA MINING MODELS FOR MINE-TO-MILL OPTIMIZATION

5.1. Case Study-A

5.2. Case Study-B

6. CONCLUSION AND RECOMMENDATIONS

6.1. Future Research

7. APPENDIX A: Impact of Different Type of Material on SAG Mill Performance and Feed Fragmentation

8. APPENDIX B: Impact of Different Type of Material on Crusher Performance and SAG Feed Fragmentation
9. APPENDIX C: Distribution of Different Type of Materials to Varying Dump Locations

10. APPENDIX D: Flowsheet of Crushing and Grinding Circuit

11. APPENDIX E: Selected SQL Queries

12. REFERENCES
List of Figures

Figure 3.1 Data-driven M2M framework ............................................................... 37
Figure 3.2 Integrated mining system ................................................................. 38
Figure 3.3 Mine-to-Mill management suite ......................................................... 39
Figure 3.4 Data model used for mine-to-mill optimization ............................... 40
Figure 3.5 ER Diagram of data enhancement ..................................................... 41
Figure 3.6 Drilling accuracy .............................................................................. 45
Figure 3.7 Explosive volumes loaded in a single shot ....................................... 46
Figure 3.8 Balanced scorecard for drilling and blasting .................................... 47
Figure 3.9 Fragmentation toolkit ...................................................................... 48
Figure 4.1 Overall view of mine (www.usgs.com) ............................................ 58
Figure 4.2 Flowsheet of copper production on mine sites ............................... 59
Figure 4.3 Effect of drillhole diameter on SAG mill feed fragmentation and amount .... 66
Figure 4.4 Drillholes and digitized free face ..................................................... 67
Figure 4.5 Effect of blasthole depth on SAG mill feed fragmentation and amount ......... 69
Figure 4.6 Effect of stemming on SAG mill feed fragmentation and amount ........... 71
Figure 4.7 Effect of water depth on SAG mill feed fragmentation and amount ............ 72
Figure 4.8 Geospatial visualization of drillhole data in RDBMS .......................... 79
Figure 5.1 Distribution of average explosive amount ........................................ 83
Figure 5.2 Distribution of fragmentation data for 7 in ...................................... 84
Figure 5.3 Decision tree model (7 in.) for Case-A ............................................. 85
Figure 5.4 Variable importance for random forest model (7 in.) for Case-A .......... 86
Figure 5.5 Variable importance for adaptive boosting model (7 in.) for Case-A ...... 87
Figure 5.6 Distribution of fragmentation data for 5 in ...................................... 88
Figure 5.7 Decision tree model (5 in.) for Case-A ............................................. 89
Figure 5.8 Variable importance for random forest (5 in.) for Case-A ................. 90
Figure 5.9 Variable importance for adaptive boosting model (5 in.) for Case-A .... 91
Figure 5.10 ROC curve of models evaluated for 7 in. for Case-A............................ 92
Figure 5.11 ROC curve of models evaluated for 5 in. for Case-A............................ 93
Figure 5.12 Decision tree model (5 in.) for Case-B................................................. 96
Figure 5.13 Variable importance for adaptive boosting model (5 in.) for Case-B...... 97
Figure 5.14 Variable importance for random forest model (5 in.) for Case-B........ 98
Figure 5.15 ROC curve of models evaluated for 5 in. for Case-B............................ 99
Figure 5.16 Decision tree model (7 in.) for Case-B................................................. 100
Figure 5.17 Variable importance for random forest model (7 in.) for Case-B........ 101
Figure 5.18 Variable importance for adaptive boosting model (7 in.) for Case-B.... 102
Figure 5.19 ROC curve of models evaluated for 7 in. for Case-B............................ 103
List of Tables

Table 2.1 Energy consumption and cost related to size reduction.......................... 27
Table 4.1 Information about tables related to fragmentation................................. 60
Table 4.2 Available variables in the data warehouse............................................. 61
Table 4.3 Burden dimension for different rock densities and explosive types (Dick et.al., 1983)................................................................................................................. 68
Table 4.4 Spacing based on hole diameter and sequence (Dick et.al., 1983).......... 69
Table 4.5 Sub-drill amount related to toe condition (Dick et.al., 1983)................. 70
Table 4.6 PMML sample of a decision tree model.................................................. 80
Table 4.7 SQL script generated from PMML of the decision tree model ............... 81
Table 5.1. Confusion Matrix for 5 in. scenario ..................................................... 94
Table 5.2 Confusion Matrix for 7 in. scenario ..................................................... 95
**List of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOP</td>
<td>Standard Operating Procedures</td>
</tr>
<tr>
<td>OLAP</td>
<td>Online Analytical Processing</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>RDBMS</td>
<td>Relational Database Management System</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio Frequency Identification</td>
</tr>
<tr>
<td>SAG</td>
<td>Semi-Autogenous Mill</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>VOD</td>
<td>Velocity of Detonation</td>
</tr>
<tr>
<td>OEM</td>
<td>Original Equipment Manufacturer</td>
</tr>
<tr>
<td>PIO</td>
<td>Process Integration And Optimization</td>
</tr>
<tr>
<td>RQD</td>
<td>Rock Quality Designation</td>
</tr>
<tr>
<td>QA/QC</td>
<td>Quality Assurance/Quality Control</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise Resource Planning</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract-Transform-Load</td>
</tr>
<tr>
<td>BI</td>
<td>Business Intelligence</td>
</tr>
<tr>
<td>ODBC</td>
<td>Open Database Connectivity</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interfaces</td>
</tr>
<tr>
<td>DCS</td>
<td>Distributed Control Systems</td>
</tr>
<tr>
<td>LIMS</td>
<td>Laboratory Information Management Systems</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>ER</td>
<td>Entity Relationship</td>
</tr>
<tr>
<td>SCADA</td>
<td>Supervisor Control And Data Acquisition</td>
</tr>
<tr>
<td>FMS</td>
<td>Fleet Management Systems</td>
</tr>
<tr>
<td>PLC</td>
<td>Programmable Logic Controller</td>
</tr>
<tr>
<td>BPI</td>
<td>Business Process Improvement</td>
</tr>
<tr>
<td>CEP</td>
<td>Complex Event Processing</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
<tr>
<td>OOB</td>
<td>Out-Of-Bag</td>
</tr>
<tr>
<td>SX/EW</td>
<td>Solvent Extraction And Electrowinning</td>
</tr>
<tr>
<td>UCS</td>
<td>Uniaxial Compressive Strength</td>
</tr>
<tr>
<td>ANFO</td>
<td>Ammonium Nitrate/Fuel Oil</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DS</td>
<td>Decision Support</td>
</tr>
<tr>
<td>AG</td>
<td>Autogenous Grinding</td>
</tr>
<tr>
<td>SRID</td>
<td>Spatial Reference Identifier</td>
</tr>
<tr>
<td>PMML</td>
<td>Predictive Model Markup Language</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>SSRS</td>
<td>SQL Server Reporting Services</td>
</tr>
</tbody>
</table>
ABSTRACT

Mine to Mill optimization is considered as a key concept for metal mining recently. Targeting operational best practices on a highly varying environment is challenging. Impact of underperformed basic operations such as drilling and blasting will sustain this inefficiency in mineral processing. Data provided for each of these operations from software and hardware utilized on field reached a level where advanced data analytics becomes applicable. In order to represent the operations as close to reality, an integrated layer of data where transactional and process based data lives is crucial. Data warehousing and data mining are alternative tools that rely on a robust data structure. Data mining utilizes the integrated data layer for pattern discovery within the data itself. Relations that are unknown for now can be investigated by data mining algorithms that rely on vast amount of data. Empirical equations that are based on a limited set of data could be improved by using data mining algorithms. The main objective of optimizing the mine to mill value chain also challenges the concept of providing real-time feedback. This research proposes a data-driven mine-to-mill framework for modern mines.
1. INTRODUCTION

The mine-to-mill process has been central to the mining industry for more than twenty years, with different approaches applied to this value chain. Various researchers have focused on different aspects of the challenging mine-to-mill problem. Mill performance is the common goal, whereas different studies focused on drilling and blasting performance through fragmentation analysis or mineral processing. Textbooks presented a distinct sequence of mining and mineral processing prior to collaborative studies in the mining industry. The mining portion of the value chain set target production figures in accordance with pre-determined cut-off values. This resulted in mineral processing having the challenging task of handling unexpected material properties in downstream operations. The importance of varying ore properties and their impact on mill performance brought to light the need to integrate objectives. Increasing interest throughout the mining industry has initiated a change in mine-to-mill understanding and prompted pioneering research in this topic.

This research proposes an innovative approach based on real-time data warehousing and the utilization of diverse data types. A framework is a set of ideas that provide guidance for a pre-determined goal. The framework described for modern mines is data-driven and covers all aspects required for an integrated operation. To find patterns and relationships between data sets, data mining was applied to a vast amount of operational data. This way, data that has a considerable impact on target fragmentation may be discovered and used as the basis for a decision support system. Operations-related data from a real-time integrated data warehouse was utilized to compare empirical approaches and models in mine-to-mill practice with actual outcomes observed at a copper mine located in Arizona. Data from the mine is anonymized to prevent any violation of confidentiality.

1.1. Research objectives and goals

The primary objective of this study is to define a framework for “data-driven mining engineering” that covers technology selection, utilization of data, change of management, and decision-making on a near real-time basis. The proposed framework focuses on re-engineering traditional mining engineering first principles that are based on equations, such as Kuz-ram, by using integrated data. This study focused on mine-to-mill optimization for an open pit copper mine, however, the
proposed framework could be implemented on other sites by using site-specific operational data. The research goals of this study are listed below.

- Mining has characteristic operations that are diverse compared to civil, oil/gas, and manufacturing industries. Best practices, standard definitions, and standard operating procedures (SOP) are commonly followed concepts in other industries whereas mining is behind in implementing these in operation. Defining a data-driven framework for modern mines that aims to optimize operations by generating data, using the technology to capture data, analyze data, and have the managerial perspective to base decision-making on data was one of the main goals of this study.

- Investment in technology became essential in the mining industry with the demand of higher production rates. In addition, low-grade orebodies force operations to utilize equipment in more efficient ways. Data used in this study originated from a corporate data warehouse of an open pit copper mine. Showing the advantages of accessing data via an integrated data warehouse was one of the goals of this research. Using Online Analytical Processing (OLAP) cubes to represent the analysis capability provided by data warehouses for mine-to-mill purposes by predicting fragmentation was one of the research objectives.

- A considerable amount of data generated by the technology utilized in surface mining can be considered geospatial. Machinery equipped with GPS is precisely tracked for both production and maintenance purposes. In this research, data gathered from drill navigation systems was transformed into a geospatial context. A nearest neighborhood algorithm was executed on a vast amount of data, taking advantage of Geospatial Structured Query Language (SQL) queries. Performance of geospatial queries and indices were optimized for faster execution times.

- Data mining is a promising technology that is implemented in various fields of research. The amount of data generated in modern mines makes the mining industry a good candidate for data mining. This study plans to use decision trees, random forests, and adaptive boosting to predict fragmentation. Drilling and blasting related data would be utilized to estimate fragmentation performance for each blasthole.
• Mine operations that have a comparably lower level of technology infrastructure available depend on manually collected data. Analysis of data by following different techniques, such as data mining to predict fragmentation relies on the quality of data. This research aims to provide necessary steps to follow through investigating available sources, cleaning, and enhancing data. The potential benefit that could be gained from utilizing data is related to the amount, content, and quality of data. However, the proposed framework provides essential stages of creating a continuous improvement initiative for mine operations either highly equipped with technology or collecting data manually.

A data-driven framework for optimizing mine-to-mill practices was introduced in this study. Multiple facets of utilizing available technology were investigated via data warehousing and data mining tools.

1.2. Research contributions

This study contributed to research in mining engineering with the framework introduced, its results, and suggestions. Each of the stated contributions is covered in either submitted journal manuscripts or draft manuscripts that will be submitted.

1) A new way of integration for drilling and blasting related data was developed by using geospatial queries in an efficient way instead of scripting complex queries in the Relational Database Management System (RDBMS). This way, data collected by drill navigation systems was enhanced and used to generate high quality operational data.

2) A unique framework for data-driven mine-to-mill optimization at modern mines was introduced by using data warehousing and data mining. The corporate data warehouse of a copper mine in Arizona was used present available data sources and their integration for mine-to-mill optimization.

3) Data mining models were trained successfully on operational data and deployed to the data warehouse to enable dynamic analysis of real-time data. Decision tree, random forest, and adaptive boosting models were used to compare variables affecting fragmentation. Both automatically and manually collected data was used to investigate parameters affecting target fragmentation.
1.3. Format of dissertation

The proposed data driven mine-to-mill framework is presented in five primary chapters following this introduction. Chapter 2 is a detailed literature survey that is grouped into four main titles of mine-to-mill practices, fragmentation models, data warehousing and data mining applications in mining industry. Chapter 3 gives details about the methodology followed to structure the data driven mine-to-mill framework. Chapter 4 describes the data mining application that was performed on a data warehouse of a copper mine. Details about data preparation, geospatial queries for enhancing drilling and blasting related data are given. Chapter 5 presents results of the data mining models trained for two cases with different scopes of data. Finally, Chapter 6 presents conclusions and recommendations for future research.
2. LITERATURE REVIEW

Metal mining is the main business activity driving research in the mine-to-mill process worldwide. The complexity of orebodies, decreasing cut-off grades, and increasing interest in utilizing data for decision making has sustained research efforts in mine-to-mill applications. Studies that have been reviewed in this chapter are primarily focused on the topics that form the background of the proposed research. This is a summary of the reviewed studies in mine-to-mill, data warehousing in the mining industry, and data mining applied to mining-related data.

2.1. Mine-to-Mill approaches and practices worldwide

Size reduction has been one of the major interests of research in mine-to-mill worldwide. Drilling and blasting is the initial focus of the operation in improving downstream processes. Cost comparisons between mine and mineral processing pointed out the importance of mill performance. Impact of the mill performance on downstream processes and vast amount of energy consumption directed research towards optimizing the mill at most of the metal mines. Common approach is to track material flow by various technologies.

Scott and Morrell (1998) pointed out that monitoring material throughout the value chain and modeling the impact of mining process on run-of-mine ore is key to improvement. The level of available technology and data limited research to approach the mine-to-mill problem from a narrower perspective. Approaches on the same concept started to differentiate with upcoming issues. Adel et.al. (2006) identified energy consumption as the main concern and concentrated to minimize energy related to size reduction of material. This also referred to drilling and blasting operations, which are proven to be the more cost effective stages of the operation that can be modified through design parameters. In this study, parameters that can be tweaked were either direct inputs for the data mining algorithm or were utilized to derive new definitions for blasthole design. The effort to focus on only minimizing energy throughout the mine-to-mill process can be considered as a rather limited approach. Various researchers investigated this problem from different perspectives.

Dance, et.al. (2011) conducted a mine to mill optimization study at Newmont ‘s Ahafo operation. Main objective was to improve plant throughput. Prior to simulation
studies, a benchmarking survey was completed as a basis of the comparison. The ore types being processed have high abrasion resistance causing a limitation of throughput at the ball mill. Material was tracked with robust RFID transducers throughout its journey from mine site to grinding mill. Drill and blast operations were revised and consistent bench preparation was initiated. Crusher performance was improved by adapting scheduled adjustments. Authors state that an 8% increase in SAG mill throughput was achieved. Amount of data gathered and utilized in mine-to-mill research is commonly limited as a robust data infrastructure for both relational and process related data is not in place. Key to efficient data utilization is integration that is covered by the fully integrated data warehouse of the copper mine used in this study. Market conditions and decreasing ore grades promoted research in increasing efficiency and implementing standards that proved themselves in manufacturing.

Meech (2000) stated that mining industry in North America is challenged by low grade and difficult to process orebodies. Recent market conditions require an overall optimization approach throughout the mine-to-mill system. Process automation, simulation, and intelligent systems for data integration are stated to be potential solutions to aid this common objective. Agent based systems utilized in manufacturing have design issues such as communication between different data sources. Common interchange languages and integration standards have a key role in data integration where different source systems talk with each other. Every technology on site provides data following different structures and requires post processing in order to be used for optimization purposes.

Herbst (2000) described mine-to-mill applications as an effort to coordinate operations between the mine and the concentrator to achieve an overall improvement in enterprise performance. It is stated that optimizing separate processes will result in a polarized sub-optimization with no combination of subsystems. Key measurements such as fragment size distribution in mine-to-mill applications are pointing out the importance of real time data gathering and analysis. Data warehousing is an efficient methodology to create semantic layers of information for further business intelligence. An ideal mine-to-mill process would enable tracking every unit amount of material throughout processes. This might be achieved by utilizing data warehouses and available technologies such as drill and blast assessment.
Morrison and Cameron (2002) reviewed available software to be used in mine-to-mill applications. Blast design and fragmentation are stated to be major actors in optimized mill throughput. Considering the mine-to-mill process as a connected chain of operations enables an overall assessment of design and operation parameters. Available image/video analysis solutions enhance the online optimization target where data being real time gains major importance. This research utilized fragmentation data provided by a high-resolution online imaging system. Parameters being tracked by available imaging technologies have the potential to be improved for better traceability.

Oghazi, et.al. (2010) studied the potential of contaminating minerals in magnetite ore for traceability in mine-to-mill applications. Mineralogy of apatite and feldspar proved that contaminated minerals survive the grinding process. Authors state that traceability enables relating process data to feed material throughout processes. However, blending is creating a complex environment and is common practice in most of the open pit metal mines. Different ore types have various mechanical behaviors in size reduction. The mill is challenged by unknown or varying hardness properties of blends.

2.2. Blast fragmentation models in mine-to-mill optimization

Performance of drilling and blasting operations is a major KPI and a valuable data source for mine-to-mill optimization. Size distribution of material after blasting or at various stages of size reduction has been the focus of research in this field. Technology is being utilized to take high-resolution images and process them by image processing algorithms. This enables evaluation of drilling, blasting, and comminution performance on real time basis. The novelty of this study is using data mining algorithms on real time integrated data to build dynamic models for mine-to-mill optimization. Empirical approaches that are based on a limited number of laboratory tests have been used in many blast fragmentation models. These models have been re-visited by various researchers, including the original authors, to improve accuracy.

Cunnigham (2005) implemented precision timing, present with electronic delay detonators in the Kuz-ram model. There were many attempts to improve this model by various researchers. Rock properties have been investigated from different
approaches to represent rock in the most realistic way. The author pointed out major risks of using this model in the wrong way and stated that this risk is caused by the ease of its application. Fragmentation models can be grouped into 2 main approaches, empirical and mechanistic. Empirical models are defined as having a common application in daily blast designs whereas mechanistic models aim for blasting results from a more broad perspective. The Kuz-ram model was not subject to a major change since its first publication (Cunnigham, 1983). The author states that current empirical models are insufficient as parameters such as rock properties and structure relative to drilling pattern, blast dimensions, timing of holes, VOD, and edge effects are not considered in detail. Another limitation of blast fragmentation models is based on current imaging technology, which has a better performance on conveyor belts compared to implementations on working benches. Laboratory scale blasting experiments were considered as an alternative but are not representative for a demonstration of the blasting mechanism and its various effects.

The adapted Kuznetsov equation is given as follows (Cunnigham, 2005),

\[ x_m = AK^{-0.8}Q^{1/6}\left(\frac{115}{RWS}\right)^{19/20} \]

where

\( x_m \) = mean particle size, cm;
\( A \) = rock factor (varying between 0.8 and 22, depending on hardness and structure) – this is a critical parameter;
\( K \) = powder factor, kg explosive per cubic meter of rock;
\( Q \) = mass of explosive in the hole, kg;
\( RWS \) = weight strength relative to ANFO, 115 being the RWS of TNT.

The Rosin-Rammler equation is as follows (Cunnigham, 2005):

\[ R_x = \exp[-0.693\left(\frac{x}{x_m}\right)^n] \]

where \( R_x \) = mass fraction retained on screen opening \( x \);
\( n \) = uniformity index, usually between 0.7 and 2.
Uniformity index is given as follows (Cunnigham, 2005):

\[
n = (2.2 - \frac{14B}{d}) \left( \sqrt{1 + \frac{S}{B}} \right) (1 - \frac{W}{B}) (\text{abs} \left( \frac{BCL - CCL}{L} \right) + 0.1)^{0.1} \frac{L}{H}
\]

where \( B \) = burden, m;
\( S \) = spacing, m;
\( d \) = hole diameter, mm;
\( W \) = standard deviation of drilling precision, m;
\( L \) = charge length, m;
\( BCL \) = bottom charge length, m;
\( CCL \) = column charge length, m;
\( H \) = bench height, m

Data utilized in the suggested model can be considered to be static, which prevents real-time optimization and on-the-fly calculations. Recent improvements in data warehousing and OEMs offering software solutions for data collection from the equipment creates the environment to integrate different data sources. This enables assessment of operations by analyzing their effect on downstream processes.

McKenzie and Adamson (2011) studied the impact of optimized delay timing for fragmentation and pointed out the need for further research in this field. Delay time per m of burden is given as the measure of delay timing, which might be utilized as an input parameter in a data mining implementation. Instead of providing raw data, enriched parameters such as scaled depth of burial or delay timing in relation with burden might result in better knowledge discovery. It is also stated that optimized delay timing for fragmentation and micro-fracturing should be considered as additional measures.

Hawkes (1998) studied the effects of improvement in drilling and blasting operations on downstream processes. The load and haul process was represented by a simulation model with data collected from a mine in Australia. It is stated that
production and cost are more sensitive to non-productive events when the shovel is not under trucked. Time gained from digging is lost when shovels have to wait for trucks. Controlled blasting that prevents over-sized blocks and boulders have major importance for excavating equipment. Data was collected manually and fit to probability distributions. Data warehouses enable live data streams from different sources to be utilized in different applications. This way, simulation results can mimic the real situation on a live basis. Digging cycle time, bucket loads, and non-productive cycles are stated to be events that are influenced by drilling and blasting significantly. Technology utilized in mining industry has increased to a level where simulation studies can be done on vast amount of data.

Herbst and Pate (1998) developed a dynamic flow sheet simulator focusing on size reduction. The spreadsheet pointed out the impact of changes in parameters such as rock properties, powder factor on mill production rates, product size distributions, and mineral liberation. The breakage event is grouped into free valuable, locked, and free gangue for simplicity. Mill objects modeled utilized rock characteristics and operational criteria to simulate the \( P_{80} \) or other trends such as percent liberation. The simulation also had additional building blocks for blasting and haulage operations, which were based on a simplified population balance - energy size relationship. Several scenarios where haulage time, powder factor, and stockpile inventory were evaluated. It is stated that a change in powder factor can be easily tracked down in product size distributions within a comparably short time. It was concluded that an ore-type database would speed up the process of generating simulation models. This point out the potential of a data warehouse utilized together with a simulation model. By having a real time data source instead of a static data repository, it would be possible to simulate real-time scenarios. This research utilized the data warehouse for mine-to-mill purposes and investigated the impact of best practices in drilling and blasting by data mining.

Dance, et.al. (2007) conducted Mine-to-Mill and Process Integration and Optimization (PIO) projects, which are claimed to result in an improvement on the concentrator throughput up to 20 percent. It is stated that PIO studies are initiated with a site visit where data about performance of the mine and processing plant is gathered for benchmarking and rock characteristics are studied for modeling purposes. Authors state that tracking material has a key role in optimization studies. In cases where fleet
management systems are not present, passive RFID tags are recommended to monitor the material movement. The tags placed in the blasthole stemming or muck pile surfaces can be tracked during their journey throughout the operation. Prior to simulation, rock structures are mapped down according to their RQD ratings and blast domains are determined. It is stated that the distribution in fine fraction after blasting has a major impact on SAG throughput being independent from the crusher operating parameters. Blast design based on assay caused a differential fragmentation with the same pattern utilized for all 3 rock domains. Variable patterns that consider different rock conditions are not practical as drill monitoring systems are not trusted as a reliable data source. It is concluded that the modified blast design with a higher cost can be seen as an investment with a considerable gain in mill throughput. This research also proved that data related to drilling and blasting is a valuable data source that can drive best practice through QA/QC.

Kanchibotla, et.al. (1998) studied the impact of blasting on downstream processes at Kargoorlie Gold mine. Blast fragmentation was modeled by a similar approach to the Kuzram model with some improvements in the effect of fines on size distribution and rock characterization. Breakage in crushing and milling is modeled by utilizing drop test results and tumbling test results. Authors utilized throughput, specific energy (kWh/t), and final grind size as SAG mill performance indicators. Suggested blast designs resulted in an increased amount of fines that reduce the SAG mill effort and a narrower distribution of critical size material that indicated less amount of oversize. Focusing on fixed parameters in mine-to-mill optimization studies might lead to sub-optimal results. Pattern discovery by data mining is taking advantage of vast amount of data and aims to expose relations that were not investigated by empirical approaches.

Pease, et.al. (1998) studied the role of process control in mine to mill effort at a lead/zinc operation. It was stated that on-stream analysis and a centralized control room supported this target. The nature of mining created a major difference by means of process control as the primary input, the ore, has varying characteristics. This natural variance cannot be closely controlled and makes manufacturing principles non-applicable in mining industry. Applying strictly defined operator rules for steady state input and output streams are not matching the needs of mineral processing industry. The suggested business plan was based on adjusting the
operation accordingly for the orebody and eliminating ore streams that will not generate a profit in the specified margin. It was concluded that a business model that adjusts the operation to the orebody is more beneficial by utilizing an adjusted version of process control. This emphasizes how crucial process related data is in mine-to-mill optimization studies. Energy consumption of crusher and mill is a KPI that is commonly used in optimization studies.

Powell and Bye (2009) handled the mine to mill approach from the energy point of view. Authors stated that major energy consuming activities in mining are material haulage and comminution. Massive low-grade deposits drive the industry to more energy consuming practices that will not fit into long-term company targets. Smart blasting is introduced as an innovative way of mining low-grade deposits and reducing the energy consumption in downstream processes. Another field of energy reduction is stated as reducing the waste that is processed in the comminution stages. Dense media and multi-stage floatation are also proposed as energy reduction strategies but the characteristics of feed material might constrain these options. Although smart blasting is stated to be promising for energy intensive size reduction, the applicability on various ore types is questionable, especially for dispersed structures. Splitting the feed flow into grade relative streams is recommended to reduce amount of energy spent to handle waste material. Upgrading the ore streams could lead to a decreased cut-off grade, which in turn would change the mine planning strategy. Ore sorting is also considered as a feasible option to prevent loss of valuable material within gangue. Larger blocks are more favorable in sorting material into potentially valuable fractions; however, close control within a multi-step procedure might be considered as an essential aspect for ore sorting. Improving equipment efficiency is expressed to be a crucial standpoint for processes such as grinding. Amount of energy used for breakage is commonly designated as an indicator of performance of mill equipment. High pressure grinding rolls might be a promising implementation in dry or low water circuits. Benchmarking processing routes on solely P_{80} is stated to be insufficient because of its weakness in handling multiple streams. It is concluded that a flexible circuit in an integrated environment will maintain highest recovery and lowest energy consumption. Drilling and blasting assessment technologies provide size distribution data on a more
frequent basis. Integrating this valuable information with other data sources will prevent approaching mine-to-mill optimization from a narrow perspective.

Scott, et.al. (1999) studied optimized blasting for mine-to-mill applications and pointed out the importance of handling optimization in this field from a multi-faceted point of view. Focusing particularly on blasting could lead to a decreased cost of excavation and other downstream processes. Energy consumption and cost of breakage is summarized as it is seen in Table 2.1 (Scott, et.al., 1999).

<table>
<thead>
<tr>
<th></th>
<th>Energy</th>
<th>Factor</th>
<th>Cost</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blasting</td>
<td>0.2 kwh/t</td>
<td>1</td>
<td>$0.15/t</td>
<td>1</td>
</tr>
<tr>
<td>Crushing</td>
<td>2 kwh/t</td>
<td>10</td>
<td>$0.75/t</td>
<td>5</td>
</tr>
<tr>
<td>Grinding</td>
<td>20 kwh/t</td>
<td>100</td>
<td>$3.75/t</td>
<td>25</td>
</tr>
</tbody>
</table>

Size distribution is the critical operating parameter of autogenous and semi-autogenous mills where the feed acts as a grinding medium. Desired distribution is defined as consisting of a larger fraction of below-grate size, not excessive amount of coarse material, and relatively little pebble size material. Reducing the effort of breakage in the mill can be considered as the main objective as the cost of improving breakage in blasting and crushing is comparably lower. Smaller top size prior to crushing makes it possible to choke feed the crusher to generate more fine material. Increased amount of fines in blasting will reduce the power consumption at the mill level.

Major aspects of size reduction considered in mine-to-mill optimization are listed as confinement, powder factor, explosive selection, blast hole diameter, burden and spacing, stemming, initiation sequence, back-break and damage, dilution, far field vibration, and air blast. Choked blasting is stated to be advantageous in case the drawbacks of trying to reduce the cost of blasting are eliminated. Another widely accepted standpoint for blast optimization is the powder factor. The common belief of utilizing powder factor as an optimization tool is claimed to focusing on loading and haulage whereas the mine-to-mill approach focuses on mill throughput. Modified blast designs for optimized mine-to-mill approaches are considered as over-
energized when powder factor is used as the basis of comparison. Reaching greater
shock energy and limiting heave energy is considered as another important concept.
This might be achieved by varying explosive densities in a column charge. Besides
the modifications suggested at blasting stage, drilling could also be improved by
various best practices.

The concept of reducing the drilling cost per ton of broken ore by increasing hole
diameter can be considered as another contradiction by disturbing the uniform
distribution of explosive energy between holes. These highly energized holes have to
be controlled by close controlled stemming to prevent loss of shock energy by
premature escape of gases. Authors suggest using variable length of stemming for
columns of higher energy zones within a blast hole. Another consideration for mine-
to-mill targeted blasting is the blast sequence. It is suggested that an even
progression of detonation front should be achieved that limits free travel distance of
each burden. Highly energized blasts would require shorter initiation sequences
compared to conventional blasts as the time required for moving the burden is
shorter.

Mobilizing the last burden becomes important when back-break and controlled
damage is considered. Reducing the explosive density or adjusting a longer initiation
at the boundary is a suggested best practice for damage control. Authors state that
dilution is the major indicator of the negative effects that might occur related to a
mine-to-mill focused blasting practice. It is concluded that KPIs such as diggability
and prevented equipment wear should be introduced into mine-to-mill and mine-to-
leach discussions as decision-making criteria.

McKee (1997) determined the linkage between fragmentation and downstream
processes in mining operations from a different perspective. Measurement of
fragmentation is stated to be difficult as the technology was not mature enough to
provide easy-to-use tools for image processing. Leaching performance of broken
rock in gold mining is stated to be another field where fragmentation has a major
impact. Structure of the heap construction and fragmentation of rock are defined as
factors controlling permeability. Rock mass properties are stated to be under-utilized
in target fragmentation calculations. Technologic advances in incorporating rock
mass properties to fragmentation models and instrumentation to estimate fine content
in muck piles are stated to be important. It is concluded that the most challenging part of a mine-to-mill project is the communication between mining and processing departments. Using scorecards and having shared KPIs are cultural enablers that require a semantic layer of data.

Gillot (2004) defines pit-to-plant as an interdisciplinary concept that benefits from mining, geology, and metallurgy to optimize comminution. Mill throughput is stated to be improved by focusing on the fines envelope adjusted by blasting parameters. Effective utilization of digital image processing via drill and blast assessment products resulted in an improved fines envelope that also increased mill throughput. The trial and error approach in blasting practices led to valuable information about the optimal parameters to be used on site. Choked blasting was preferred to reduce dilution, however, the conventional ratio of 80% emulsion and 20% ANFO resulted in excess amount of heave energy and undesired back-break. This supports the idea that optimization studies in mine-to-mill increase the mining cost while reducing downstream processing costs. The culture barrier to increase a cost related to drilling and blasting has to be handled by setting clear and shared goals throughout the operation.

Marek (1991) clarifies the importance of the mine and processing plant staff working in a harmony. Any goals set by the managerial level have to be canalized through different departments to reach operational managers and technical employees. Communication between processing and mining departments is the key point in any optimization study on mine-to-mill. Both sides should be well aware of each other’s problems and concerns about production. Data warehousing and business intelligence are promising investments for mining companies aiming to closely control processes for optimization purposes.

Jankovic and Valery (2002) conducted a feasibility study on mine-to-mill for a gold mine in Australia. Size distribution of stream throughout the comminution circuit is monitored by digital image processing and used as simulation inputs. It is stated that blasting with an objective to create a wider range of fine portion for mill feed would enable an increase in throughput and relax the crushing circuit. By applying Bond energy calculations on real data, it was pointed out that the decrease in the ball mill’s energy consumption would outweigh the increase in energy consumption from finer
crushing. Real-time data could provide required information on a continuous basis so that optimization would not be a static and one time effort.

A recent research project conducted in Sweden by Ouchterlony (2013) stated that recent development in imaging technology and related classification algorithms has changed the way fragmentation analysis is conducted on site. Various applications like zoom merge or fines corrections have enhanced image analysis systems. Another point that is outlined is about potential energy savings within the mine-to-mill stream. Blasting is considered to be the cheapest method to reduce size of rock; therefore, investing in a higher specific charge at this stage should be preferred instead of consuming higher energy at crushers or mills. Close control fragmentation supported with a solid strategy for mineral processing can be considered as a leading actor for the mine to mill process.

Besides parameters that can be controlled, analysis and interpretation of fragmentation via systems that analyze and interpret fragmentation can be considered as another major stage in mine-to-mill approach. An increase in the amount of undersize that bypasses crushing stages or decrease in the feed size of the crushers will be advantageous for downstream processes such as grinding. Improvement in grinding also heavily depends on level of micro fracturing as these structures have a higher possibility to survive in crushing. Material is considered to improve hardness in each stage of size reduction and the presence of blast induced fractures decrease during comminution.

There are certain indicators that could be used to measure the effect of fragmentation as the number of maintenance or down events such as bridging or plugging at the crusher caused by oversize material. Cycle times might also be pointing out performance changes related to fragmentation. Although grinding is commonly stated to be the most important cost driver, fragmentation also affects cost of loading, haulage, secondary crushing, and other processes. Therefore, fragmentation size distribution plays an important role in all stages of mine-to-mill applications. The cost breakdown of size reduction given by Herbst and Pate (2003) is based on cost of wear parts and points out that 77% could be related to grinding whereas only 1% is allocated to explosives fracturing. Inefficiency in grinding could also be related to the
energy lost as heat. Decreasing the Bond work index (1959) is a common measure that is utilized in cost reduction for mine-to-mill implementation.

\[
W = 10 \cdot W_i \left[ \frac{1}{\sqrt{X_{P80}}} - \frac{1}{\sqrt{X_{F80}}} \right]
\]

where \( W = \) Work input (kWh/ton)

\( W_i = \) Work index for the specific rock type (kWh/ton)

\( X_{P80} = 80\% \) passing size of the product (µm)

\( X_{F80} = 80\% \) passing size of the feed (µm)

Recent development in access to operational data through data warehousing introduces potential of real-time data. An integrated layer of data could feed parameters such as \( P_{80} \) or \( F_{80} \) continuously with data that might lead to new perspectives on various processes of mine-to-mill. This approach should not be seen as a focus on specific processes as any improvement in upstream processes would result in better loading and haulage performance, reduced maintenance events, and related cost. However, interrelations between stages such as crushing and grinding might be more complex than expected. Material that bypasses primary crushing could be a goal for blast optimization while it would increase the top size of the coarser side of the size distribution. Such crusher product might adversely affect Mill performance. Therefore, data available from all processes in mine-to-mill approach should be considered as potential inputs for improvement.

**2.3. Data warehousing for mining related data**

Data warehousing can be defined as a collection of databases in an integrated way where the time perspective can be historical or real-time (Inmon, 2005). Its benefits in management and decision-making have been proven in various industries. Mining is a highly complex production field where technology is utilized heavily on equipment. Mobile equipment that is part of the fleet management system provides relational data whereas plant equipment generates non-relational process data. Integration of these unique data sources for business intelligence on a real-time basis is a challenge.
Barr and Cook (2008) discussed strategic alignment in the mining industry. It was pointed out that data gathering should be done top-down whereas analysis of alternative strategies should be throughout the organization. Communication is seen as the key to success in continuous improvement. The integrated data warehouse utilized for the mine-to-mill study in this research has the capability to provide in-depth analysis to all levels of operation. Balanced scorecards are an example of cultural change in mines towards acting on data.

Hunter (2011) emphasized the importance of delivering actionable information to key personnel. Spreadsheets that are commonly used as personal databases are stated as inaccurate data handling tools. The risk of keeping valuable information in an environment that can get inaccessible after a certain file size is another drawback of spreadsheets used in continuous improvement. Outsourcing development tasks is stated to be time consuming which shows the necessity of easy-to-use tools. A robust data warehouse as used in optimizing the mine-to-mill value chain of a copper mine provided a flexible environment for browsing integrated data. This way all personnel are equipped with necessary information on a real-time basis.

Bascur and Soudek (2010) discussed the potential of historians as the integration level. Asset frameworks were defined as elements of the system where calculation could be embedded and logic defined. Main focus is to transform process data into actionable items. Extending equipment availability and reducing operating cost are other objectives of implementing real-time integration of mining and metallurgical data. Efficient use of water and energy in mining industry has a crucial role and can be related to the process side of the operation primarily. Having process data integrated with other types of operational data is key to optimizing the mine-to-mill concept as a whole. The integrated data warehouse used for this optimization study took advantage of both historians and relational databases. Keeping data in their native environment and integrating them in a semantic layer after cleaning and normalization supports the idea of getting to integrated data as real-time as possible.

Barr and Cook (2007) highlighted the necessity of real-time performance measurement that enables critical business intelligence. It is stated that data visibility becomes more important at the operator level where it could impact production directly. Main objective in implementing real-time performance measures is explained
as getting the right KPIs to the right people at the right time. As it was the case for the data warehouse utilized in this research, a single source of correct information throughout the company aids business visibility and decision-making. Integrating data with ERP systems and accounting software provides on-demand performance of operation. This research utilized a fully integrated data warehouse to both analyze and conduct data mining on real-time operational data.

MacDonald (2013) pointed to the inefficient utilization of data in mining companies. KPIs that measure performance of employees in different departments are kept within their native environment. Sharing information and building common goals towards a global objective is recommended for better decision-making. The increase in throughput claimed as 8% by implementing a process automation system shows how real-time data aids setting priorities. Operating in a seamless data environment as a close loop of real-time data is foreseen as the future of mines and mills of the future. The real-time data warehouse utilized in this optimization study is a big step in achieving this optimal industry goal.

Danckert and Masterman (2011) presented another example of centralized data used in orebody knowledge. OLAP and Extract-Transform-Load (ETL), commonly referred in data warehousing studies, were used to access detailed information in block models. Existing block models and newly developed ones used for reserve estimation and mine planning are stored in a centralized data repository. Such kind of centralized data is utilized in an operation room for integrated planning and maximizing system efficiency. Such innovative implementations drive change of management culture from individual operations towards a single integrated system. Centralizing data and managing it from a control room is different from integrating different source systems.

Smit (2002) discussed benefits of an integrated asset management system at a metallurgical plant. Transactional data provided by accounting systems give insight to financial measures but have limited capability to contribute value drivers. Authors stated that high-speed business environment would require delivery of quality BI to the managers in future. Performance measurement and strategic decisions based on data are foreseen as key actors in managing mines in the most effective way.
Benito and Dessureault (2008) pointed out the fact that data used in block models got more complex and also increased in size. Pit optimization is conventionally performed on economic figures and geological data such as grade. Current trend is to utilize as much data as possible in an integrated environment. The well-known Lerchs and Grossmann graph algorithm was revisited by using relational databases. Coding was completed in SQL environment and results were visualized via Open Database Connectivity (ODBC) connections. The major advantage of working in a relational environment is the flexibility of browsing data via Pivot Tables. It is possible to drill-down block model data to determine short-term plans based on detailed data. Data warehousing is an essential part of advanced implementations on mining related data such as OLAP cubes and data mining.

Kahraman and Dessureault (2011) performed another implementation of an integrated data warehouse by creating a real-time adherence to mine plan tool. The idea of maximizing production while ensuring quality revealed the necessity of a production-tracking environment. GUIs designed from real-time integrated data are stated to be aiding staff to perform their work with least amount of effort. Peer-pressure driving higher effort, time saving, and simplified reporting are some of the many advantages of a data warehouse that are highlighted. This study was a major part of a decision support system designed for a remote control room.

Tenorio and Dessureault (2011) designed a remote control room that highlights various applications of an integrated data warehouse for business intelligence purposes. One of the main components of the decision support system is a simulator that reproduces the mine site and performance of essential equipment. A real-time adherence to mine plan tool is providing hourly and shift based production data. The fleet cost management tool is based on activity based costing and aids in tracking operating cost real-time. Operators were evaluated to achieve minimum blending error with maximum utilization using all available tools in the control room. It is concluded that real-time integrated data becomes more and more crucial in decision-making as time is the major constraint in operation.

2.4. Data mining for knowledge discovery in mining related data

Mine-to-mill can be considered as one of the most challenging applications based on the variety and amount of data available. Discovering relations between different data
sources has been studied by a narrow group of researchers. As stated by Han et.al. (2006) data mining is defined as extracting or “mining” knowledge from large amounts of data. The integrated data warehouse used in this research is a robust source of clean and real time operational data.

Benito and Dessureault (2013) discussed an advanced implementation of data warehousing and data mining for estimating final pit limit. Haulage cost is major driver in financial analysis and related data mainly lives in accounting and fleet management systems. Integrated data from these source systems was fed to a decision trees algorithm to investigate strongest relation to haul time. Incorporating haul time and cycle time data enhanced the block model used for final pit estimation. Size of final pit showed differences for optimized cycle time and haul time. It was concluded that the proposed idea of using data mining in pit optimization could be improved by updating input parameters with historical data.

Buxton et.al. (2014) studied the potential of sensor data for real time mine optimization. It is stated that common utilization of sensor data is in pre-concentration and online characterization. Eliminating waste material from ore as early as possible can be considered as an optimization goal that can benefit from sensor data. Compliance to the plan is another implementation where real time data could be utilized to support decision-making. Statistical inference methods applied to link sensor data with other information like rock properties are expected to become more popular. Data mining could be an alternative or supporting tool to statistical approaches when applied to vast amount of real time data.

Krishnan (2012) pointed out the challenges of current mining and mineral processing activities and how they could be handled with the aid of IT platforms. Data integration between distributed control systems (DCS), enterprise resource planning (ERP), and laboratory information management systems (LIMS) at mineral processing plants is stated as a well-known example at manufacturing industry. Amount of data generated makes data mining a viable option to predict accuracy in handling outliers or missing values in data related to plant performance.

Hales et.al. (2009) discussed how plant performance could be improved by artificial intelligence and expert systems. Business rules are stated to be inefficient due to
their static nature. Analyzing data with least amount of human skills is seen as target to promote automated data handling. Data mining is stated to be applicable for rule induction, which has a major impact on expert systems. Current improvement in fields like cloud computing and parallel processing make data mining more applicable by providing software and hardware support on demand.

Bogunovic et.al. (2009) discussed how data related to energy consumption in surface coal mining could be handled by modern IT components. An integrated data environment system was built to integrate different systems although dragline related data was gathered manually from the screen. Data mining was utilized to source raw data from the DCS. Main focus was to point out hotspots for energy consumption. The data environment coupled with a GUI gave access of crucial data to key personnel in a flexible and convenient way.
3. METHODOLOGY OF DATA-DRIVEN MINE-TO-MILL OPTIMIZATION FRAMEWORK

3.1. Data-driven Mine-to-Mill Optimization Framework

The core idea behind this study is to introduce a data driven mine-to-mill (M2M) framework for modern mines. The foundation of this concept is data that requires investment in technology to generate, collect, and analyze it. This stage also involves the managerial perspective, which recognizes the crucial role of utilizing data. By using data warehousing tools, continuous improvement departments can provide information in a flexible environment based on integrated data. Once data is cleaned and integrated, advanced analysis methods, such as data mining, can be performed on a real-time basis. The uniqueness of this study is the way data was integrated across different levels of granularity, enhanced by geospatial queries, and analyzed via data mining algorithms that are deployed back into the same environment through the data warehouse. This closed loop approach of data is represented in Figure 3.1.

![Data-driven M2M framework](image)

**Figure 3.1 Data-driven M2M framework**

The contemporary M2M framework changes the conventional way of generating information and acting on it. Data is the main resource that the framework is built on to optimize the mine-to-mill value chain. However, there are other important concepts that are essential for the proposed framework. Figure 3.2 represents the integrated mining system that incorporates the data-driven M2M framework.
The integrated mining system consists of all major drivers related to the data to action concept, which are platform, process, and people. Platform can be defined as the IT-backbone that covers sources, communications, analytics, and infrastructure. As it is closely related to the hardware and software used in the system, it is discussed more commonly among different levels of management. A common mistake of management is that investing in technology is considered as sufficient to build an integrated mining system. In recent years, this management mindset recognized the major role of the process stage that is the core of using the data. Infrastructures that are built on data, as part of the framework in this study, are not sufficient to optimize a value chain, such as mining. The mine-to-mill oriented framework requires an integrated layer of data that is part of the platform. People that are capable of using this data to generate knowledge are highly valuable assets that support the proposed framework. Organizational strategy and change management are concepts that are performed by people that are trained on using data and form a corporate culture for continuous improvement initiatives.

The suggested framework supports a contemporary mine-to-mill practice in which information is a shared asset between departments. A data warehouse provides a single source of reliable and integrated data. Data collected from various systems is
cleaned and enhanced through geospatial queries. The results of advanced analysis methods, such as data mining, are fed back to the data warehouse for further analysis. This framework is built on a data-driven management perspective and supports the utilization of data in daily operations by using different tools as shown in Figure 3.3.

Figure 3.3 Mine-to-Mill management suite

An integrated data warehouse creates an environment suitable to developing tools and performing analysis on operational data. OLAP cubes, dashboards, and scorecards are some of the business intelligence implementations that could support managers, supervisors, and operators. Data warehousing is a powerful system that improves itself through key personnel using it. Any suggestion or correction to data can be implemented in a single source of data, so that all custom tools built on the same data layer will benefit. Therefore, the data itself is crucial for a data warehouse. Mining is a multi-faceted industry that generates various types of data. Building tools to manage each stage of operation and calculate performance metrics on the data requires a detailed characterization of the data sets prior to integration. Mine-to-mill is a complex value chain that can be optimized through data warehousing and data mining, as in this research. The data model created to enhance existing data sets and prepare inputs for data mining algorithms is given in Figure 3.4.
The fact table used as the basis of the data model is an integration of drill monitoring systems, reports of explosive loading, fleet management systems, and the drill and blast assessment system. Granularity of this table is every photo taken by the drill and blast assessment software. It was aimed to use the lowest level of granularity so that all available data sources could be integrated in a meaningful way. The integration point was chosen as a single blasthole that is defined uniquely throughout different systems. Drill navigation systems record data for each blasthole whereas explosive loading is commonly reported on a scale of multiple blastholes defined as a shot. However, using data related to drilling and blasting for a group of blastholes reduces the quality of available data by assigning a single value to multiple records. For instance, the data warehouse used in this study provided values for hole diameter, stemming, and other drilling related parameters that indicated no variance. Data mining was part of the proposed framework and input variables were preferably not unique values of limited amount. Data mining models perform better when trained on data that has a certain level of variance. Therefore, derivation of innovative parameters related to mine-to-mill was an objective of this study.

Geospatial queries were used to find closest holes to each blasthole and determine a distance parameter that is unique. Using static parameters together with the varying distance provided the basics of enhancing data. Finding the closest holes to each unique blasthole and populating all associated data, such as diameter, stemming,
and average penetration rate created the derived table. The new fact table is an enhanced version of the original fact table including close blastholes, distances, and other values associated to them. The entity relationship (ER) diagram given in Figure 3.5 summarizes the closed cycle of data enhancement.

![ER Diagram of data enhancement](image)

The nearest neighborhood algorithm implemented in the RDBMS as a stored procedure was used to find close blastholes. Cursor logic was developed to go through the raw fact table row-by-row and rank blastholes by their distance to each other. The geospatial fact table is based on the geospatial column formed by using X and Y coordinates of blastholes. Geospatial indexing was also applied on the table to increase query performance. Once the distance and ranks of blastholes were found, incorporating them with the static parameters generated new measures that had variance used in data warehousing implementations and data mining.

### 3.2. Mine-to-mill Optimization Based on Data Warehousing

Mine-to-mill optimization studies are commonly based on limited data sets and approach the problem from narrow perspectives. This research, on the other hand, developed a novel method of handling available data sources and implemented data mining algorithms on a vast amount of operational data. Different systems utilized for
fleet management, fragmentation assessment, drill navigation, and plant performance provided valuable data to define the value stream, from drilling to the product of mineral processing. An integrated data warehouse of an open pit copper operation provided the context for the research. The following chapter introduces the characteristics of data related to the mining industry and its integration for business intelligence applications. These applications enhance raw data and provide an easy-to-use environment to analyze and report a vast amount of data on a real-time basis.

3.2.1. Integration of relational and process data through data warehousing

Optimization in the mining industry can make a greater impact by focusing on achieving improvements across the multiple processes in the production stream to facilitate improved quality, throughput, and ultimately profit. Data is a critical enabler to continuous improvement initiatives. Data of different forms and levels of detail are found throughout industrial processes. For example, in open pit mines, equipment is typically monitored through GPS-based technologies that produce relational data, essentially large data sets saved in the form of tables with associated relationships among the tables. In mineral processing plants, floor production information is viewed through Supervisor Control and Data Acquisition (SCADA) systems whose data is largely analog, stored through data loggers, and more sophisticatedly, processed by data historians. The integration of relational in-pit data with plant data can help improve the sustainability and regularity of a key aspect of metal and minerals production, often called “mine-to-mill”.

3.2.1.1. Characteristics of data in mining industry

The scope and definition of data in the mining industry changed over time with technological improvements in software and hardware products. The utilization of data increased as mine management perspectives changed, requiring increasingly reliable and real-time data. The most commonly used data types in mining are relational and process related data.

The primary source of relational data in mining are fleet management systems (FMS), which became an industry standard after their introduction in the 1980s. Utilizing GPS technology in FMS revolutionized the way in which mines were managed. Mobile equipment, which plays a major role in production, is tracked by
high-precision GPS and sensors for optimal production and equipment health. These are event type relational data sources and define operational cycles in detail. Each basic motion of the equipment, such as spotting or dumping for haul trucks, is recorded in a spatially meaningful way. Drills are also monitored in a similar manner by gathering data associated with their position on the drill plan.

In addition to time and location, other information related to production can also be considered relational data. The number of holes drilled, tons dumped, and other measures are commonly utilized for analysis and decision-making. Decreasing grades for metal mines, increasing mineral demand, and changes in operational perspectives steer the mining industry to a more data-driven management style. This results in a need to integrate different types of mining related data. A common data integration source is transaction relational data, which may be provided by enterprise systems, mine planning databases, or safety systems.

Another primary type of data for the mining industry is process-related data, which has a completely different structure and characteristics than relational data sources. Process control systems, environmental monitoring systems, and SCADA systems are the main sources of process data. Crushers, belt conveyors, mills, and other critical components of the value stream are equipped with sensors or tags. These tags collect data in a more primitive way than relational data sources. However, the data frequency is much higher which requires a different IT strategy to process the data.

Mineral processing related data typically consists of a measurement and a timestamp. Power consumption, tons per hour, and run time are basic measures that are tracked. As the frequency of process data is much higher than that of relational data, process data from sources such as Programmable Logic Controllers (PLCs) are handled by hardware components called data historians. Historians are capable of rapidly reading and writing data records. Process related data is small in size, but its frequency creates a vast amount of data. Isolating the data to certain departments limits the potential for creating value for the whole value chain.

Data integration is crucial for exploiting the technology investment made at every stage of production. Analyzing data and creating metrics to aid decision-making or
performance measurement on an integrated data environment has various benefits. The semantic data layer utilized in this research has the capability to handle both relational and process data on a real-time basis. Reports, OLAP cubes, and dashboards are some of the applications based on data warehouses. More advanced implementations, such as simulation or data mining, are enabled by an integrated data warehouse.

3.2.2. Business process improvement for drilling and blasting practices

Business Process Improvement (BPI) initiatives can often provide greater impact when focused on achieving improvements across multiple processes in the production stream to facilitate improved quality, throughput, and ultimately profit. Data is a critical enabler of BPI initiatives. Data of different forms and level of detail are found throughout industrial processes. Mining equipment is commonly monitored through GPS-based technologies that produce relational data, essentially large data sets saving data in the form of tables with associated relationships to each other. In mineral processing plants, floor production information is viewed through Supervisor Control and Data Acquisition (SCADA) systems whose data is largely analog, stored through data loggers, and more sophisticatedly, data historians. The integration of relational in-pit data with plant data can help improve the sustainability and regularity of mine-to-mill, a key BPI area in metal and minerals production. The framework used in this study introduces practical steps, best practices, and leading software tools necessary to implement a sustainable mine-to-mill at a copper mine.

The integrated data warehouse utilized in this research provided business improvement in numerous aspects of the operation. Drilling and blasting-related performance enhancements are the building blocks of a sophisticated mine-to-mill strategy. Mining is a heavily equipment-dependent industry, in which both manpower and machine are crucial. Therefore, both of these assets should be analyzed in detail. The technology and data available in modern mines creates the necessity of real-time business intelligence implementations.

GPS data sourced from drill navigation systems provide the exact locations of drilled blastholes on site. Drillhole locations are subject to controversy, as there are different sources for this data. Drillhole locations are first designed in mine planning software and then marked on site by surveying. Drill plans are uploaded to drilling machines
and operators also consider survey marks. As a result, the drill navigation system collects GPS data while operating. The difference in these data sets indicates drilling accuracy, which can be visualized as shown in Figure 3.6 and used as a measure.

![Drilling Accuracy Diagram](image)

**Figure 3.6 Drilling accuracy**

As seen in Figure 3.6 the difference between the locations of drillholes as designed and as drilled is a major performance indicator for drilling operations. Visualizing data and providing an easy-to-use environment to drill down in the data layers is key to business improvement. Such accuracy metrics can also be considered as the basis for Quality Assurance/Quality Control (QA/QC) practices in drilling. Utilization of this data enabled a corrective action that benefited supervisors and operators. Discrepancy in drilling accuracy decreased and maintained its target value through daily data analysis. Another QA/QC benefit that data warehousing provided was related to explosive loading.

The amount of explosive loaded in blastholes has to be closely controlled for target fragmentation, environmental impacts, and stability concerns. Data related to explosive volume loaded was extracted from the contractor’s daily report. The variation in explosive volume loaded in neighboring blastholes indicates potential
inefficiencies. Figure 3.7 shows that a single shot consisting of blastholes with the same design parameters produces inconsistencies in the amount of explosive.

![Graph showing explosive volumes loaded in a single shot](image)

**Figure 3.7 Explosive volumes loaded in a single shot**

This unbalanced distribution of explosives could have adverse effects such as fly rock, air shock, vibration, and over-size rocks. Non-uniform energy distribution can be considered as one of the major reasons of missing target fragmentation. The performance of downstream processes as digging, loading, haulage, and crushing is prone to be affected by uncontrolled explosive amounts. The integrated data warehouse utilized in this study provided a flexible environment to analyze blasting related data. Reporting on inefficiencies such as these, led to process improvement in drilling and blasting. Reports also fill the role of cultural enablers.

Balanced scorecards can be defined as hybrid representations of both financial and operational performance measures. Kaplan and Norton (1992) discussed the importance of information systems in drilling down into detailed data when required. The integrated data warehouse utilized in this research is an example of a highly granular real-time data environment for a modern mine. The business intelligence
potential of scorecards is based on goals and measures. Managers, who have a vision of reaching a certain level of quality, see higher level goals, whereas supervisors, operators, and other operations staff need more detailed measures on which they can act on. Figure 3.8 represents a drilling and blasting scorecard.

![Scorecard Image]

Figure 3.8 Balanced scorecard for drilling and blasting

As seen in the scorecard, predetermined key performance indicators are used to summarize daily operations for the drill and blast manager. Internal communication could be improved by utilizing such scorecards as a single, reliable source of information. Goals shared among different departments create valuable discussion points and support continuous improvement. One of the major contributions of such BI implementations as scorecards is the self-improvement these data enable. As seen in the drill accuracy measure, missing design files that should have been uploaded to the drill navigation system from the mine planning software indicates an issue. The manager has the right tools to catch such problems, justify facts based on actual data, and take action on them by communicating with related staff. A fragmentation toolkit as shown in Figure 3.9 supports managers by providing an integrated view of the data.
Data from drill navigation systems, an explosive loading system, drill and blast assessment software, and a fleet management system was integrated in the data warehouse. Penetration rate, an indicator of rock strength and drill performance provides more value when analyzed together with the amount of explosive loaded. A single shot of multiple blastholes can be used to represent the effect of upstream processes on fragmentation and digging performance. Although this BI implementation generates valuable information, the amount and frequency of data in a modern mine requires a more sophisticated approach.

### 3.3. Mine-to-Mill Optimization through Data Mining

Data mining was introduced in the late 1980s by the database community, which successfully developed infrastructure for handling a vast amount of data (Piatetsky, et.al., 1991). The collaboration of statisticians and computer scientists, especially researchers in the field of machine learning, led to the increased popularity of data mining and its application in various industries. Together with growing agile concepts in software engineering, data mining activities started to follow basic project management principles. Face-to-face communication between parties, preferably with the participation of an industry expert, is a key concept that guides a data mining project to success.

Data mining algorithms grew more sophisticated as the amount, complexity, and frequency of collected data advanced. Concepts like “Big Data (Cox and Ellsworth, 1997)” drove the need to consider new tools, such as Complex Event Processing (CEP) (Luckham and Frasca, 1998), NoSQL (Strozzi, 2010), and real-time data mining. In addition to commercial packages for data mining, there are sophisticated
open source software tools developed by researchers in this field. R (Ihaka and Gentleman, 1996) is an open source statistical toolkit that provides data mining functionality via a native coding language. It is also possible to utilize the power of R through the use of Graphical User Interfaces (GUI) such as Rattle (Wills, 2009), without deep knowledge of coding R scripts, although there is migration support between environments.

Besides building and executing data mining models, it is of critical importance to perform descriptive statistical analysis on input data. Understanding the general trends, skewness, and distribution of data aids the evaluation of a designed data mining model. Predictions or knowledge discovery studies that provide results in favor of an already known fact might be viewed as accurate but would have limited value. Once the model is built and executed, evaluating it becomes crucial. Sample data is separated from the whole data set to train, test, and validate the model. Performing evaluation of the model by using training data leads to decisive results. As training data is used in building the model, it is prone to bias; therefore, testing and validation data sets should be used for accuracy testing. A validation data set is utilized in the assessment of the model developed based on training data. Tweaking available parameters improves model performance prior to comparing it with testing data.

3.3.1. Data cleaning and modeling in Rattle

Data is the key point in any data mining project, as its quality, amount, availability, and other characteristics determine the time and effort required for analysis. The objective of creating knowledge from data highly depends on the characteristics of the data itself. Data collected by different systems in a similar way result in an integration challenge. Matching records on a common basis, such as time or location, requires detailed analysis, programming, and knowledge of mining technology. The integrated data warehouse utilized in this study is a highly sophisticated environment that provides operational data cleaned and integrated on a near real-time basis.

A table in data warehousing terms is known as a “data set” in Rattle. Rows are designated as observations, whereas columns are known as variables. Identifiers are unique variables that distinguish each observation from another. Primary keys, which would not serve as an input to data mining projects if they do not represent any
knowledge, are commonly used in data warehousing. Accessing data sources requires a certain skill set applied to ETL, databases, and RDBMS. Rattle is capable of utilizing ODBC to connect to MS Excel, MS SQL Server, MS Access, and other data repositories. This provides the flexibility to manipulate data while importing it into R via SQL query support.

Data sets from SPSS and SAS can also be loaded into the R library to be accessed by Rattle. Using XML is an efficient way to transfer over data from web sites. Copying and pasting data from a spreadsheet or another source is an option for importing data into R. Sampling the data set is initially preferred, as the computing time and use of limited data for model building can optimize performance. Data can be extended to cover larger populations once the model is developed, although the class imbalance issue remains. The major disadvantage of sampling is the probability of losing outliers from a sample that could be of interest when running a data mining project. Default sampling is set to be 70%, 15%, and 15% for training, validation, and testing respectively. Such settings and seed values, used to initiate random number generation, can be tweaked to optimize the model. Different approaches that use seed values as a control parameter while building the model are common in data mining projects.

Data are categorized as input, target, or identifier by considering the number of distinct values for that variable. Exploring data prior to any modeling effort is essential to assess data quality before using it for data mining purposes. Descriptive statistics support the preliminary stages of data mining, where the data set is explored in detail. Parameters such as skewness, kurtosis, and missing values count can determine how data is distributed (Ramachandran and Tsokos, 2009). Skewness greater than one indicates that data spreads towards one direction of the mean based on positive or negative skewness values. Kurtosis defines the shape of the distribution based on peak values. The impact of these parameters is greater for traditional approaches than for recent data mining algorithms that consider a variety of factors. It is possible to create box plots, histograms, cumulative distribution curves, and Benford’s Law plots for numeric data, whereas bar plots, dot plots, and mosaic plots are prepared for categorical data.
R provides a wide variety of plotting options, due to the extent of its developer community. Such graphical output becomes more valuable when the relationships between variables are analyzed. Correlation analysis is one of the fundamental steps of data analysis for investigation of multiple variables together. A correlation coefficient close to either -1 or 1 represents a strong relation between the two parameters where a very strong correlation might indicate that the variables are not independent. This condition has major significance for the performance of some data mining algorithms. Therefore, expertise in defining variables and their inter-relation should be one of the preliminary steps of data mining projects.

A correlation plot represents a relation both by shape and color. A straight line converts into a circle with declining correlation. The orientation of the shape indicates whether the correlation is negative or positive. Additional scale elements use shades of red for negative correlation and blue for positive correlation. Data with missing values can also be used for correlation analysis, although the count of each variable should be taken into consideration to prevent any wrong conclusions. Plots that are static provide insight to the data, although utilizing dynamic graphics offers more potential for investigation. GGobi (Land and Swayne, 2001) is an extensive visualization package for R that can be utilized for dynamic plots.

Data cleaning is a key process in any data mining project, regardless of the source of data. This research utilized an integrated data warehouse that pre-processed multiple raw data sources and stored them in a normalized structure. This significantly reduced the effort required to cleaning the data prior to building models. Different aspects of cleaning data, such as misspellings, missing values, and outliers must be understood in detail. Rescaling is a type of transformation that aims to prevent any of the variables suppressing the weight associated with another variable during clustering. Normalization is a common method applied to get all variables into the same range of values. This controls the difference between numeric values of varying magnitude.

Recentering, in which the difference between each variable and the mean value is divided by the standard deviation, is another useful technique. Recentering rescales data values. Similarly, median and median absolute deviation can be utilized for data frames in which outliers are targeted. In the case of a skewed distribution, a
logarithmic transformation can be applied to variables to obtain values closer to each other. Missing values, another common issue of data frames, have to be investigated in detail prior to modeling. The distinction between having a value of zero and other reasons of missing data, such as a malfunctioning sensor, should be made if imputation will be performed. Zero values might fill gaps in the data frame; the mean may be used if there is a normal distribution, or the median if may be used if there is a significantly skewed distribution.

Changing the type of data may also improve its visualization for data analysis, although this should only be performed after considering the impact. Binning separates numeric values into four default categories that can be tweaked while using equal count, k-means, or equal width options. On the other hand, some cases require that categorical variables be represented as numeric values. This type of transformation is preferred for k-means clustering and other distance-based models, together with some type of regression models. Data cleaning is a stage of data mining that is commonly performed.

3.3.2. Data mining models

Data mining focuses on building models that can be classified into two main categories, descriptive and predictive (Jiawei and Kamber, 2006). The main objective is to represent a real problem, analyze related data, and develop a systematic and robust approach to either regenerate the problem with different inputs or discover how certain aspects are related to each other. Such model building techniques as clustering, association, and pattern discovery are descriptive. Descriptive models are used when a target variable is not required to expand current understanding of a concept, as forecasting is not the focus of descriptive analysis. Classification and regression models are common methodologies utilized in descriptive analysis. Classification models aim to associate new observations into correct classes, which are related to the target variable values. Regression models, on the other hand, are more quantitative, intended to predict a numeric value.

Cluster analysis is a powerful tool used to divide a data frame into smaller data sets that are easier to manage (Hand et. al, 2001). The main objective of clustering is to indicate the difference between clusters that are formed by similar data. A well-known representation in clustering is k-means, which utilizes a search heuristic to build a
model. A random group of k clusters is first represented by a group of variable mean values, and then the distance of a variable to the collection of k vectors of that mean is calculated in an iterative manner. The search for the model is finalized when the data points are not moving between clusters. There are more sophisticated clustering models in which entropy-weighted variable selection is used to enhance the quality for data frameworks with many dimensions. Multi-dimensional data is also commonly analyzed using association rules.

Association analysis, which aims to find relations between parameters, has proved itself to be a powerful tool for large transactional data sets (Williams, 2011). The common occurrence of items together might be a potential association rule. Various algorithms were developed to reduce the computational demand for search heuristics. The performance of association models can be optimized using such parameters as the support that defines the threshold for frequency. The outcome of this analysis is a set of rules with their probabilities of occurrence. Confidence is another parameter that could be adjusted to find only strong relations within the data frame. Another common method is decision trees, which were developed for machine learning.

Decision trees originate from a root node, and its leaves are split by tests (Hand, et al, 2001). Branching of the tree is driven by rules that define how partitions will be formed. Translation of decision trees into rules is performed by means of coding. Utilizing a mathematical representation of rules is common during implementation of decision criteria in programming languages, SQL Server in this research. Data partitioned into sections is easier to process and enable the splitting of decision trees until no more knowledge can be discovered. There is a trade-off caused by selecting a variable that might not be the best option and using it until the tree reaches a certain level of detail. A definition of entropy is used to measure performance of potential splits and the amount of information needed to improve the model.

Irregularities in the data frame are evaluated with respect to the target variable to determine whether enough information is present to categorize the data for prediction. Entropy or the amount of information required for classification in decision trees is given by the equation (Williams, 2011).
\[ info(D) = -p \log_2(p) - n \log_2(n) \]

Data frames that are naturally split into equally possible observations indicate higher entropy; in other words, more information is needed to assign data to categories. In this approach, each split possibility in the decision tree is evaluated based on the resulting entropy and the information gained by the split. Recursive splitting of the tree and choosing the best variable continues until dividing the dataset does not improve the results or there is insufficient data. A conclusive tree identifies the best variables and their optimal splits, which can be used by decision support tools. This study aimed to use drilling and blasting related variables for the target variable of fragmentation to discover knowledge about the mine-to-mill value chain.

The decision tree provides graphical output and detailed information about each split in the form of textual output. Details about the node, where it was split, how many entities are present, what the resulting performance was, the default classification, and the distribution classification are examples of interpretations that can be made once the tree is built. Each node can be investigated in detail to track how the model reacted to missing values or to observe the improvement trend. Performance of the model is characterized by such parameters as complexity and cross error, both for categorical and numeric target variables.

Categorical variables are used in classification models, whereas numeric variables may be modeled by using analysis of variance (ANOVA) to create regression trees. The minimum number of observations for splitting nodes can be modified to control tree growth. Building a tree with no limitations may be preferable to see where performance parameters like complexity reach their optimum level. This allows parameters to be tweaked to improve results. Although decision trees are computationally fast and compact, there is a possibility that the best available tree cannot be constructed. Random forests and boosted decision tree algorithms are techniques that were developed to handle this drawback of naïve decision trees.

Random forests essentially aim to construct multiple decision trees that are equally good and to join them in a conclusive, single model (Breiman, 2001). Randomness associated with the selection of variables and observations provides an efficient
structure to handle outliers, missing values, and overfitting. Overfitting is a concept seen in predictive data mining algorithms that defines the case in which a training data set is perfectly classified by compromising the test error. Fully-constructed, smaller decision trees, which might individually be over-trained, build up random forests. However, the combined result is not overfit to the training data set.

Adaptive boosting algorithm builds decision trees on a random basis by bagging. Bootstrap aggregation, also known as bagging, is a concept based on sampling observations into smaller sets and using those sets to build decision trees. Multiple subsets of random data are split into training and validation sets for each decision tree. Improvements in performance when compared to decision trees result from the variation between smaller subsets, which provides the flexibility to choose the best decisions among them. Another benefit of utilizing small-scale decision trees is its computational cost, as a small subset of variables selected is used to build decision trees without any pruning. Performance of the resulting model can be evaluated in various ways.

3.3.3. Evaluation of data mining models

Out-of-bag (OOB) estimation uses data not used in training decision trees, providing an unbiased error estimate. It represents how the developed model will behave when new data is introduced. Another way of evaluating model performance is the confusion matrix, which summarizes the actual and estimated values of the target variable. In case the training dataset is dominated by a certain value or category, it is possible to balance the error in estimation by modifying sampling ratios. This way, the model will use the determined amount of sample for building the trees and provide a better estimation. The majority of false positive and false negative cases should be evaluated in detail prior to adjusting the sampling parameters of the model.

Measuring the importance of variables is another major result of random forest models, as success in estimating the target variable is associated with the variables used to train the decision trees. This allows the model to be re-trained by using pre-determined variables for better computational efficiency. The relative comparison of variables based on their accuracy measure is commonly used to modify the training strategy of decision trees. Similarly, the number of trees included in the forest can be
evaluated by considering the error rate, so that a threshold value can be discovered, beyond which no significant improvement is seen.

Another sophisticated model builder is the boosting algorithm, which is considered one of the best off-the-shelf classifiers (Hastie et al., 2001). Similar to random forests, boosting utilizes various learning algorithms that might not be the best performing alternative when used in isolation. Weights are associated with observations and prioritized according to their relative performance in modeling. The primary difference between this technique and random forests is that trees are built in a self-developing fashion so that observations that were not classified properly in the previous tree are boosted in the next tree. The initial weights of observations are changed each time a tree is built and evaluated. Observations that cause errors will be boosted until the model accuracy reaches the desired level.

Evaluation of models can be done in various ways. The most important part of any evaluation is the data used for scoring. Data is split into partitions prior to model building so that an unbiased set of observations are preserved for evaluation purposes. Training data that is utilized to construct the model is a questionable choice for testing the performance of a data mining algorithm. A rather biased picture will be seen when the model is evaluated using observations from the training data set. Validation data provides a better estimate of the model’s performance when compared to training data. An evaluation conducted during model building and fine-tuning should be based on separate validation data. Once the final model is obtained, using validation data that was utilized during the model’s development will introduce bias. Testing partition, which uses new data that has not previously been presented to the algorithm, provides a realistic assessment of the model’s performance. Evaluating a model using new data provides a better indication of its performance in future.

Error rates are straightforward measures based on comparing actual values with predicted values. Classifying the errors made by the model in accordance with hypothesis testing is beneficial when a misclassification case is practically significant. Weights can be assigned to certain types of errors so that the model is forced to reduce the error rate for these cases. This is a matter of defining whether a false
positive or a false negative case has more severe impacts on realistic cases. The final stage of a data mining implementation is deployment, during which the model is exported and re-used with new data.
4. APPLICATION OF BUSINESS INTELLIGENCE AND DATA MINING

4.1. Information about the mine
An open pit copper mine located in Arizona owns the integrated data warehouse used in this study. It covers multiple pit locations, ore crushing and milling stages, and floatation facilities besides leaching operations. The mine is a conventional open pit truck-shovel operation and consists of smaller mine phases with varying rock conditions. Daily production of the mine reaches approximately 250,000 tons operating in along with a concentrator and a solvent extraction and electrowinning (SX/EW) plant. Concentrate produced on site is sent to a smelter by railway where copper anodes are manufactured. These anodes are then sent to a refinery whereas the SX/EW plant handles leach ore. An overall view of the mine site is given in Figure 4.1.

Figure 4.1 Overall view of mine (www.usgs.com)
Drilling and blasting are crucial stages of the operation for this mine as mill and leach material require different levels of fragmentation. This leads to the effort of starting production on site where drilling is considered as the initial and a crucial stage of the value chain. Drill monitoring systems are used to track daily drilling operations and support QA/QC targets. Explosive loading is handled by a contractor, which is also closely monitored by daily reports. The fleet management systems equipped on haul trucks and loading equipment provides detailed data about the production cycle on site. Similarly, the historian and process control system manages process-related data generated by equipment, such as crushers, grinding mills, flotation cells, and other machinery used in mineral processing. The mine is operating in association with other mine sites of the same company and general structure of production is given in the flowsheet in Figure 4.2.

![Flowsheet of copper production on mine sites](image.png)

**Figure 4.2 Flowsheet of copper production on mine sites**

The data warehouse used in this study was provided by Mine-3. Mine-to-mill value chain starts with drilling and blasting followed by loading by shovels. Haul trucks are the primary equipment used in haulage and dump material to stockpiles or the crusher. This stage is where mining meets mineral processing, defined as
concentrator in this case and where related data changes structure. A flowsheet illustrating the high-level structure of the concentrator is given in Appendix D. Sulfide, classified as high-grade, is fed to the crusher and then to the SAG mill whereas low-grade sulfide and oxides are sent to heap leach pads. Leaching is out of scope for this study; however, it should be studied in future research aiming to prepare material on site by drilling and blasting for better leach performance.

4.2. Information about data sources available in the data warehouse
The real-time, integrated data warehouse used in this study has a structure that provides a semantic layer of operational data for a modern mine. The primary sources of data cover the areas of accounting, fleet management, equipment health, process control, drill navigation, drill and blast assessment, geoscientific data, and safety. Transactional data generated by accounting and fleet management systems has large volume and high frequency. Process control data are not transactional; however, they are more frequent and are collected via small-size packages. Differences of these data types create challenges during data integration. The data warehouse used in this study has a robust infrastructure to scale up in case a new data source is added. Some of the main tables in the data warehouse are presented, together with their row counts, in Table 4.1.

Table 4.1 Information about tables related to fragmentation

<table>
<thead>
<tr>
<th>System</th>
<th>Row Count</th>
<th>Start Date</th>
<th>End Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Control</td>
<td>569,677,056</td>
<td>6/26/2008</td>
<td>4/12/2015</td>
</tr>
<tr>
<td>Fleet Management (Haulage)</td>
<td>1,472,141</td>
<td>1/1/2010</td>
<td>4/12/2015</td>
</tr>
<tr>
<td>Fleet Management (Loading)</td>
<td>1,470,033</td>
<td>1/1/2010</td>
<td>4/12/2015</td>
</tr>
<tr>
<td>Fragmentation</td>
<td>587,080</td>
<td>2/1/2011</td>
<td>4/12/2015</td>
</tr>
<tr>
<td>DnB Assessment</td>
<td>425,861</td>
<td>2/1/2013</td>
<td>4/12/2015</td>
</tr>
<tr>
<td>Drill Navigation System (current)</td>
<td>43,421</td>
<td>5/9/2012</td>
<td>4/12/2015</td>
</tr>
</tbody>
</table>

Data collected by these systems are staged in the data warehouse prior to any integration. Explosive consumption and blasthole design geometry data, manually prepared by field staff, are rather static when compared to real-time data sources such as drill navigation systems. The granularity of the manually collected data is also less than that of other data sources in the data warehouse. However, this data is
valuable in validating and enhancing the data, as its content is not captured by any other system on site.

The vast amount of data available in the data warehouse provides input parameters for data mining purposes. Knowledge discovery is one of the main targets of data mining so that all data sources were investigated for potential usage. Issues were found in some systems that recorded erroneous data for certain time ranges that were handled in the data cleaning stage by filtering out. Different systems that are implemented on site performing the same action also made it challenging to integrate data in some cases. Drilling can be given as an example where different drill monitoring systems were used on site over the years; however, only recent data was found to be a potential input for data mining. Table 4.2 summarizes the available variables related to mine-to-mill in the data warehouse.

Table 4.2 Available variables in the data warehouse

<table>
<thead>
<tr>
<th>Drilling</th>
<th>Drill &amp; Blast Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilled hole count</td>
<td>Avg. of Top size</td>
</tr>
<tr>
<td>Avg. Pen. Rate</td>
<td>Avg. Pictures taken per truck load</td>
</tr>
<tr>
<td>Number of redrilled holes</td>
<td>Avg. of F\textsubscript{80}, F\textsubscript{60}, F\textsubscript{20}, etc.</td>
</tr>
<tr>
<td>Horizontal offset</td>
<td>Avg. of conveyed material on belts</td>
</tr>
<tr>
<td>Toe offset in X, Y, and Z</td>
<td>Hole depth</td>
</tr>
<tr>
<td>Avg. duration to drill a hole</td>
<td>Stemming height</td>
</tr>
<tr>
<td>Collar offset in X, Y, and Z</td>
<td>Hole diameter</td>
</tr>
<tr>
<td># of holes with no design file</td>
<td>Explosives load height</td>
</tr>
<tr>
<td>Hole depth</td>
<td>Water level</td>
</tr>
<tr>
<td>Actual avg. drilling rate</td>
<td>Explosives amount</td>
</tr>
<tr>
<td>Collar coordinates in X, Y, Z</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explosives loading (automatic)</th>
<th>Explosives loading (manual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. subdrill per shot</td>
<td>Avg. SAG mill fresh feed</td>
</tr>
<tr>
<td>Avg. plugged holes per shot</td>
<td>Avg. SAG mill total motor power</td>
</tr>
<tr>
<td>Avg. diameter per hole</td>
<td>Avg. SAG mill trunnion bearing prs.</td>
</tr>
<tr>
<td>Avg. spacing per hole</td>
<td>Avg. SAG mill sound</td>
</tr>
<tr>
<td>Avg. depth per hole</td>
<td>Avg. SAG mill density</td>
</tr>
<tr>
<td>Avg. burden per hole</td>
<td>Avg. SAG mill speed</td>
</tr>
<tr>
<td>Avg. stemming per hole</td>
<td>Avg. pile height</td>
</tr>
<tr>
<td>Avg. powder factor per hole</td>
<td>Avg. Crusher motor amps</td>
</tr>
<tr>
<td>Avg. depth per hole</td>
<td>Avg. Crusher mantle position</td>
</tr>
<tr>
<td>Rock density</td>
<td>Avg. Crusher circulating load</td>
</tr>
<tr>
<td></td>
<td>Reagent consumption, etc.</td>
</tr>
</tbody>
</table>
Main objective of using a data warehouse for mine-to-mill optimization purposes was to utilize integrated data in a timespan during which all systems had data. This resulted in a high quality data set covering a shorter time range. Another reason of limiting the data set is the lack of material tracking between the mine operation and mineral processing site. Material cannot be traced by data once haul trucks dump it although it is equipped with a fleet management system. The destination logged as crusher is insufficient as material could be moved to different stockpiles which will undergo dozer pushing with equipment that is not tracked by GPS. This issue created a gap between mining related data and process type data from mineral processing. Missing data, especially related to rock mass, is another problem of the data warehouse.

The mine that provided the data warehouse for this study lacks a geotechnical department. Therefore, data defining rock strength such as uniaxial compressive strength (UCS) is missing. Rock types and mineralogy were reviewed from different sources in literature; however, studies about slope stability and underground openings were not preferred. No laboratory test results were available as the mine that provided the data warehouse had no assigned staff for geotechnical studies. As part of the suggested framework, alternatives to deterministic approaches were aimed. Average penetration rate was used to classify blastholes to be drilled through varying strength, so that the impact of rock was partially represented. These issues are part of future research and planned to be handled in upcoming studies.

4.3. Business Intelligence applications on mining related data

Business Intelligence (BI) has evolved together with technology over time, and new implementations will continue to evolve and sustain its major role in various industries. The boom of big data is a phenomenon that is currently changing the scope of BI. Although the scope and technology of BI have evolved, its objectives have remained the same objectives. Luhn (1958) defined the basic expectations of a business intelligence system as providing information based on user demand. This became common practice in many industrial applications, including mining.

This study focuses on mine-to-mill applications for an open pit copper mine. Due to its interdisciplinary nature, different types of data were utilized in an integrated environment. The real-time data warehouse built for the copper mine was first used
to analyze data via OLAP cubes and then to perform data mining. Various parameters related to drilling, blasting, and fragmentation were studied in the scope of analyzing mining related to through OLAP cubes.

4.3.1. Geological conditions

Researchers have studied rocks as engineering materials starting as early as 1960s, with increasing interest in civil and mining applications (Hoek, 2007). Drilling and blasting are operations in which rock is the primary focus in the initial stages of the mine-to-mill value chain. Investigating the geological conditions and properties of rock is a pre-requisite for a successful blasting operation. Effort in achieving a target fragmentation starts with understanding the rock and its behavior during blasting operations. As defined by Latham and Lu (1999), blastability can be defined as the ease of fragmentation of a rock as a result of a blasting operation. Geologic properties that affect blastability are covered broadly within the scope of this study, as the main focus is on parameters that can be modified during operations.

4.3.2. Structure

Rock, composed of various types of minerals with different grain sizes, is a complex material to handle. Although drilling provides information about rock properties, it is not feasible to predict fragmentation solely based on this data. The structure of rock can be explained by parameters related to its orientation, such as strike and dip; discontinuities, such as joints and faults; and composition related terms, such as grain size.

Strike and dip can be considered interpretations of the orientation of rock. There are noteworthy concepts in blasthole design related to dip and strike. A common practice is to create a blast face that is aligned with the dip, aiming for a flat bottom, which is advantageous for digging purposes. Blasting against dip might result in a certain extent of backbreak, which is observed behind the last row of blastholes. Similar to dip, blasting against strike could lead to backbreak and an irregular floor that will impact subsequent blasts. Shovels are key mobile equipment in production. Operational data of shovel monitoring systems provide precise GPS locations and sensor readings that can be interpreted as downtime. Therefore, any condition that
will increase its production cycle or create additional risk on its components should be tracked from reliable data sources so it can be prevented.

Blasthole design that considers joint systems could be used to advantage in blasting. Keeping the blasting face parallel to the major joint system may be beneficial for moving a rock mass towards the free face. Joint angles and the orientation of joints that cross a certain orientation are likely to cause backbreak and unexpected fly rock conditions. Hard rock formations that are naturally sized by joints are favorable for blasting. Similarly, mineral composition and matrix structure building up the rock are unmodifiable parameters that impact blasting. Powder factor and amount of explosive required to blast a unit measure of rock, might be tweaked when reliable information is available. Test blasts and drill monitoring systems provide important data about mineral composition and how it behaves under certain conditions. Weak zones within the formation are prone to cause problems, such as air blast and fly rock. Explosive charged in weaker zones or cracks might result in an irregular face.

4.3.3. Rock strength

Rock strength is a major property that limits the efficiency of blasting. Fragmentation is commonly defined by mechanisms that follow a sequence of breakage events. Gases released as a result of explosive initiation aim to separate a rock volume internally, which is related to tensile strength. The compressive strength of rocks is stated to be higher than its tensile strength in many cases. Shock waves that reach a blasthole wall create a crushed zone due to exceeding the compressive strength. Breakage based on tensile failures occurs and triggers radial cracking until the shockwave intensity decreases. Laboratory measurements of representative rock samples are essential to understand the mechanical properties of the location. However, strength is not a parameter that can be adjusted for mine-to-mill purposes.

4.3.4. Rock density

Rock density is a key factor for displacing the blasted rock volume, especially in cast blasting. Blasting is known as the most cost-effective process to move material. To achieve target displacement, powder factor has to be adjusted based on the rock density. Similar to rock strength, density is a property that is identified by laboratory tests. Drilling operations provide a continuous data feed via drill monitoring systems;
however, interpretation of difficulty in drilling might be misleading. Based on its resilience, hard rock might be harder to drill but easier to blast under certain conditions. Understanding the properties of rock is beneficial for estimating the outcomes of a blasting operation.

4.3.5. Blasthole design

Geological conditions, which cannot be controlled, are the major drivers of fragmentation. The mechanical properties of rock and geologic discontinuities are implemented as constants in empirical approaches. This chapter will consider parameters that are commonly utilized in blasthole design based on formulas from previous researchers in this field.

4.3.6. Blasthole diameter

Blasthole diameter is primarily related to the drilling equipment on site. However, performance of blasting is commonly predicted by using empirical equations based on diameter, burden, depth, and other parameters. Critical diameter is defined as the limiting dimension of the blasthole, below which the charged explosive has the potential to fail to be initiated (Rustan, 1998). Explosive manufacturers design product containers and cartridges in accordance with the critical diameters of different types of explosives.

Bulk explosives that are charged with special trucks perform better in larger diameter blastholes. Decreasing the blasthole diameter might negatively affect detonation velocity, which defines the rate of the reaction triggered by initiation. This property has a crucial role in blasting, as the reaction that will create extensive amount of energy and gas release might fail if the detonation velocity is too low. Changing the diameter requires adjustments in the drilling pattern and might change the related cost. Smaller diameter holes and tighter patterns increase cost but have the potential to prevent boulders.

The following data is from a copper mine in Arizona that initiated its mine-to-mill program in 2013. The integrated data warehouse supplies real-time data from every stage of the value chain. However, certain assumptions are still necessary, as the material dumped at the crushers is not closely tracked by the available technology. Crushed material has several potential routes, such as dozer pushing to a stockpile,
delivery to a concentrator without further grinding, or feed to the Semi-autogenous grinding mill (SAG) by the surge bin. The historian stores process data is integrated with the data warehouse. Tags are used to gather data from equipment utilized in the mineral processing stage. Surge bin level is monitored by both tags and laser measurements for material tracking. However, the composition of the material stored in the surge bin, fed to the SAG mill, is not clear. Therefore, upstream process related data could only be associated with the material stream by making certain assumptions for a particular mine. Engineers on site stated that the crushed material reaches the SAG mill in 40 - 45 minutes. The effect of blasthole diameter on the SAG mill fresh feed could be analyzed in a more precise way if a time model is created for crushed material. This is a subject that will be considered in future research opportunities.

Blasthole diameter is rarely changed for the same material type. Different diameters may be preferred in production or overburden blasts for varying rock strength or target fragmentation. Consideration of diameter as a single parameter does not affect either average fragmentation size or the average amount of SAG mill fresh feed, as shown in Figure 4.3.

Figure 4.3 Effect of drillhole diameter on SAG mill feed fragmentation and amount
Changing the diameter within the same shot plan is not common practice. The mine used for the case study prefers pre-determined and fixed blasthole designs. The variation in fragmentation and tonnage for SAG mill feed cannot be explained by diameter alone. This indicates that diameter might have an impact if it is analyzed in combination with other parameters. Empirical approaches might not always overlap with real data. Modifying a parameter such as drillhole diameter during operations for research purposes presents another challenge.

### 4.3.7. Burden

Burden is a design parameter that is the basis of well-known blasthole design approaches. Theoretically, burden is defined as the shortest vertical distance to the free face from the center of charge (Rustan, 1998). However, burden calculations are prone to uncertainty when multiple free faces are present. Drill monitoring systems provide a high-precision GPS heartbeat to determine hole locations. Burden is not a measure that is tracked by current technology; however, the real-time data warehouse can be used for this purpose, too. Actual burden calculations could be achieved by utilizing the digitized string of the bench design from the mine planning software, if available, and coordinates from the drill monitoring system as depicted in Figure 4.4.

![Figure 4.4 Drillholes and digitized free face](image-url)
The main aim of the proposed dynamic calculation is to reduce uncertainty through the use of high-precision GPS data. The present study utilized geographical and geometrical data for spatial calculations in MS SQL Server. Shortest distance calculations for every blasthole were integrated in the data warehouse. This required integration and back-end coding in Transact-SQL (T-SQL) so that the real-time data warehouse was enhanced to become a complete solution for the mine-to-mill purposes.

In addition to the dynamic determination of burden, several practical approaches were taken based on previous mining experience. A well-known fact is that having too small burden might result in an undesired amount of forward throw, whereas too large a burden would result in deficient fragmentation. Dick et.al (1983) and other researchers clarified the practical burden based on rock density, as shown in Table 4.3.

### Table 4.3 Burden dimension for different rock densities and explosive types (Dick *et.al.*, 1983)

<table>
<thead>
<tr>
<th>Rock Density</th>
<th>Explosive Type</th>
<th>ANFO</th>
<th>Slurry / Emulsion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light rock (2.2 g/cc)</td>
<td></td>
<td>28 x diameter</td>
<td>33 x diameter</td>
</tr>
<tr>
<td>Medium rock (2.7 g/cc)</td>
<td></td>
<td>25 x diameter</td>
<td>30 x diameter</td>
</tr>
<tr>
<td>Dense rock (3.2 g/cc)</td>
<td></td>
<td>23 x diameter</td>
<td>27 x diameter</td>
</tr>
</tbody>
</table>

#### 4.3.8. Spacing

Spacing is another design parameter that is commonly held constant during a drilling operation. It has a crucial role in blasting practice, as charged explosives might lose their combined impact by being separated by too much spacing. Conversely, blastholes that are too close have the risk of causing undesired vibration levels or reduced explosive impact. Practical rules that are defined for spacing are based on burden. Bender (1999) stated that spacing should be one or two times the burden, and it might be changed based on the diameter and blasting sequence, as indicated in Table 4.4.
Table 4.4 Spacing based on hole diameter and sequence (Dick et al., 1983)

<table>
<thead>
<tr>
<th>Spacing type</th>
<th>Spacing</th>
</tr>
</thead>
<tbody>
<tr>
<td>holes shot instantly by row</td>
<td>1.8 - 2.0 x burden</td>
</tr>
<tr>
<td>large diameter holes shot sequentially</td>
<td>1.2 - 1.5 x burden</td>
</tr>
<tr>
<td>small diameter holes shot sequentially</td>
<td>1.5 - 1.8 x burden</td>
</tr>
</tbody>
</table>

The logic and back-end coding in the data warehouse that determines burden was also used to calculate spacing. As it is preferable for spacing to be held constant, dynamic logic is unnecessary.

4.3.9. Bench height

Bench height is controlled by the mine planning and equipment selection processes. Dimensions of the digging equipment, in this case a shovel, are considered together with certain safety precautions to determine the bench height. This parameter might have an impact on fragmentation; however, it is not modified during drilling and blasting operations. Drillhole depth, a more commonly utilized parameter, is equal to the sum of bench height and sub-drill amount. Figure 4.5 represents the impact of hole depth on SAG mill feed fragmentation and amount.

Figure 4.5 Effect of blasthole depth on SAG mill feed fragmentation and amount
The similar trend between hole depth and amount of SAG mill feed indicates that deeper blastholes charged in a correct way might produce more material. However, the relation between fragmentation and drillhole depth could not be explained by using depth as a single parameter. This implies that a more complex combination of parameters is required to enlighten the real impact on fragmentation.

### 4.3.10. Sub-drilling

Sub-drilling has an impact on dig performance and shovel metrics. An insufficient amount of sub-drilling might cause irregular and challenging digging conditions for the shovel that also risk the health of the machine. Besides the time and resources consumed in drilling, excessive sub-drilling might risk bench stability; a smooth bench surface is desired. Bender (1999) related sub-drill amount to burden, as shown in Table 4.5, and stated that it is commonly 0.1 to 0.5 times the burden.

<table>
<thead>
<tr>
<th>Toe condition</th>
<th>Sub-drill</th>
</tr>
</thead>
<tbody>
<tr>
<td>flat bedding plane at toe</td>
<td>0.0 - 0.1 x burden</td>
</tr>
<tr>
<td>relatively easy toe</td>
<td>0.1 - 0.2 x burden</td>
</tr>
<tr>
<td>medium toe</td>
<td>0.2 - 0.4 x burden</td>
</tr>
<tr>
<td>difficult toe with vertical bedding</td>
<td>0.5 x burden</td>
</tr>
</tbody>
</table>

### 4.3.11. Stemming

Stemming is subject to a multi-perspective approach, as both the length and type of material used are major factors in blasting. It is a common practice to use drill cuttings as stemming material. Although this may be preferable, varying shapes of drill cutting might also reduce the quality of stemming. A major objective is to sustain gas pressure that is generated as a product of detonation so that it leads to proper fragmentation. The water condition of blastholes is a factor in determining stemming length. Different stemming lengths and materials are suggested for various conditions, such as having shorter stemming for stone chips in dry holes. Reported
Stemming heights from the integrated data warehouse and their impact on SAG mill feed and amount are summarized in Figure 4.6.

Figure 4.6 Effect of stemming on SAG mill feed fragmentation and amount

Stemming, recorded manually on site, has a certain level of uncertainty. Human error is one of the major drawbacks that cause inaccuracies in the stemming data. As stemming length was recorded as being constant, its impact on fragmentation cannot be analyzed without combining it with other parameters in this study. This highlights the necessity of implementing more sophisticated tools, such as data mining, to recover the relationships and levels of impacts.

4.3.12. Water depth

Presence of water is a well-known challenge in blasting, especially when ANFO is consumed as blasting agent. Although it is not a preferred condition, considering the impact of water depth on fragmentation would emphasize that the importance of following standard operating procedures in blasting. Figure 4.7 represents the water amount present in the blastholes and its impact on SAG mill feed.
Figure 4.7 Effect of water depth on SAG mill feed fragmentation and amount

Analysis of water depth as a single parameter for fragmentation does not indicate a similar trend. More insight into its effect might be found when it is used in data mining algorithms as part of the proposed framework in this study. The effects of various parameters on mill feed in this chapter are based on actual data from a copper mine. The SAG mill is considered the most important activity in the value chain. The impact of different types of mill will be discussed in the next chapter. Rod mills are more conventional grinding equipment in comparison to SAG mills, which are recently more popular at copper mines.

4.3.13. Material type

The case study presented in this research is a copper mine that produces ore from four different locations. Each pit location has a different type of rock domain and related characteristics that affect SAG mill performance and feed fragmentation, as evident in Appendix A.

The impact of blending material with different properties has no significant impact on SAG mill run time. However, the increasing trend in average amount of SAG feed indicates that mining multiple locations simultaneously aids fresh feed production. A homogeneous feed is primarily desirable due to the adverse impact of blending on
recovery. If harder material, such as DS location, is dominating SAG mill fresh feed, a certain level of fragmentation is sustained. Different materials change the $F_{80}$ value, which indicates the necessity of analyzing blasting operations for different pit locations. This adds complexity to the concept of mine-to-mill, as making suggestions for target fragmentation should also consider rock type as a major factor.

Crusher mantle position and power drawn during the crushing stage are measures that give insight to processes upstream of the SAG mill. Any optimization of blasting will first be observed at the crushing stage and then in subsequent processes. Figure in the Appendix B represents the relationship between crusher operational parameters and SAG mill feed properties.

Figure in the Appendix B also clearly indicates that different types of material with varying hardness have no effect on the motor amps drawn by the crusher. The mantle position of the crusher indicates the changes in top size and the presence of boulders. Boulders are unexpectedly large-dimension material that might require additional re-handling before being fed to crushing circuits. The number of boulders is a measure of the success of the blasting operation. An increasing trend of mantle position for DS material type might reflect inefficiencies in blasting, whereas blends from softer rocks occasionally result in narrower settings.

### 4.3.14. Mill type

Rod mills and ball mills in a closed circuit with crushing were considered as conventional for a long time. Development in Autogenous Grinding (AG) and SAG mills are commonly used in copper mines nowadays, and the product size distribution is different from that of rod mills. The grinding media in AG and SAG mills is either entirely or partially the material itself, so the product size distribution is distinctive from other circuits. Improvement in key components of SAG mills and mines that operate on multiple pit locations at the same time led to the popularity of this type of grinding equipment.

Multi-pit operations are common in today’s mine operations, as meeting production requirements is a continuous challenge. This results in blending different material and necessitates close control of blasting, as mill performance is sensitive to changes in material type. The mid-sized mine used in the scope of this study has four main pit
locations. Appendix C shows the amount of material from different pit locations that are dumped at different locations.

The numerous dump locations and production fields with different material properties represent the complexity of current copper mining operations. Therefore, blast fragmentation optimization studies should consider material type as a major factor, even though these geological parameters cannot be modified.

SAG mills stand out from conventional circuits with rod mills by the higher amount of fine generation in its product stream. Rod mills commonly have a narrow size distribution band with smaller amounts of fines in product. Another major difference in using SAG mills is the requirement for a certain level of rock hardness as the grinding media is partially the grinded material itself. This also explains the higher ratio of production from the DS zone, which contained harder material, in the case study. Blasting fragmentation would be affected by the higher amount of explosive required to blast the harder rock material.

Rod mills tend to grind most all of the coarse-size material fed, but the product size would not be fine as compared to that of SAG mills. On the contrary, SAG mills grind a lower percentage of coarse material but produce finer product. The fundamental differences of the grinding mechanisms of these two mill types define the fragmentation of feed properties. When a rod mill is used instead of a SAG mill with the objective of a fine product size, fine feed with a considerable homogeneity is required. Rods used in the mill are efficient in reducing the size of coarse material that comes into contact with the grinding media.

Rod mills are well known for their reduction of the slime that has to be removed from the feed prior to floatation. Blasting fragmentation for a rod mill with a target fine size might require more fine production, which would guide the blasthole design. Increasing the powder factor might be an option to increase the amount of fines in blasting. Modifying the grinding circuit by focusing solely on the mill might not be the best option. Classifiers and crushers should also be investigated to match best performance in the milling stage.

One conclusion is that it is not easy to define the blast fragmentation requirements for different types of mills. There are many components that affect the performance of
the mill, including crushers and classifiers. However, it can be stated that a rod mill would tend to require finer feed material, which could be achieved by increasing powder factor, using stronger explosives, and designing narrower blasthole patterns by reducing spacing and blasthole diameter. This issue was also addressed by Kanchibolta et.al. (1998). High quality data and expert opinion are the basis of research to handle such complex modifications in any stage of the mine-to-mill concept.

4.4. Preparation of mining related data for data mining applications
Business intelligence and data mining applications utilize data as a resource. Therefore, the quality of data used in analysis, reporting, or other applications are of major importance. All types of data, including the commonly relational or process based data in the mining industry, are prone to errors in the various stages of their life cycle. Data collected manually or automatically can be affected by human error or device malfunctions. Hardware such as GPS, sensors, PLCs, or even servers can lower data quality.

This study leverages the existence of a copper mine’s integrated data warehouse to optimize the end-to-end mine-to-mill process. Conventional data warehousing implements staging as one of the fundamental steps where raw data is sourced via ETL processes and cleaned prior to storage in tables. This data manipulation includes consistency checks in different dimensions, such as time or location. Missing values and outliers are also preferably handled in a systematic way to maintain a desired level of consistency.

Data related to mine-to-mill practices were sourced from different systems and integrated for business intelligence purposes. Starting from the initial stage that involves the drill-monitoring system, location-based drilling metrics were extracted and transformed in the data warehouse. Drillhole related data is classified by unique names and enhanced by high-precision GPS coordinates. Performance of drilling equipment is sourced from on-board hardware whereas design parameters such as depth and spacing are extracted from design files uploaded from mine planning systems. Fleet management systems provide equipment related data on a real-time basis with precise location. Loading equipment that operates on blasted material is
tracked with sensors and high-precision GPS to achieve high availability and utilization.

Similarly haul trucks that transport blasted material from a loading location to stockpiles or the crusher are closely tracked by the fleet management system. Location and timestamps are used as common integration points to join data from different sources. Data integration is key to represent the value chain in real-time, supported by a single source of reliable data. The granularity of data varies between a blasthole or shot in drilling, a bucket load of shovels, and a truck dump at the crusher throughout the mine-to-mill value chain. Data integration of these different data sources can cover all aspects of the mine-to-mill process until the final product is achieved, wherever material can be associated to data throughout the process. Data validation in this study was performed by rules derived from previous best practices applied to raw data to improve data quality. This stage included strategy development for handling missing records and outliers.

Data gathered by the drill and blast assessment system was cleaned by removing missing values and performing a consistency check for outliers. Fragmentation values were classified as fine and coarse as the purpose of the data mining application was to discover relationships instead of deriving an equation. Threshold values of 5 inches and 7 inches were used to separate data into its categories. These threshold values were selected based on discussions with drill and blast supervisors and the mine manager. Department scorecards have goals defined for fragmentation at loading equipment and feed material of the mill. Converting fragmentation values into a categorical variable for data mining provided the flexibility of applying all available models for training. Another reason of not using the target variable as a numeric value is the evolution of data mining practices. The software used for data mining in this study has up-to-date packages to estimate variables that are in text format. Similarly, average penetration rate that is sourced from the drill navigation system was redefined as rock strength. This improved the accuracy and partially filled the data gap related to rock mechanics related parameters.

Fact tables for building OLAP cubes were utilized as the primary data source throughout this study. To incorporate existing know-how in different stages of the mine-to-mill process, data were manipulated to enhance the information that can be
gained out of it. Data mining algorithms were implemented in various domains to enhance the meaning of the data used as input. Therefore, the algorithm might perform better with a guided approach that represents the relationship between input variables for the mining industry.

Unique blasthole names and data associated to each blasthole such as diameter, depth, explosive amount, water amount, average penetration rate, and $F_{80}$ are stored in a single table. A stored procedure that implements a nearest neighborhood algorithm was applied to this table. Stored procedures are groups of SQL statements that are developed to execute commands on a systematic basis. The number of holes to be searched and the maximum distance from the reference hole are adjustable parameters of the stored procedure. The ten closest blastholes, if present, within a range of 50 units were identified for each unique blasthole.

The GPS coordinates of blastholes were used to convert data into a geospatial meaningful data set for further analysis. The advantage of using geospatial data in an RDBMS in this study was the availability of functions for calculations and the geospatial index, which could boost performance. Data related to the closest blastholes were joined to the same fact table by calculating the minimum, maximum, and average values. In this manner, the groups of closest blastholes were inserted into the fact table, both increasing the amount of data available and enhancing data.

**4.5. Derivation of innovative parameters for data mining**

Drilling and blasting data, which have major influence on fragmentation, dominate the behavior of the mine-to-mill value chain. In addition, drilling and blasting are upstream processes that have numerous operating parameters that can be tweaked for optimization of the overall process. Design parameters such as burden, spacing, and diameter are basic variables of existing blasting and fragmentation theory. Availability of technology and data in today’s mining operations removes the constraint of deterministic approaches.

Utilization of operational data collected on site is essential for continuous improvement. Therefore, data analysis to optimize mining operations should be part of managing every modern mine, which uses an integrated and flexible data environment. Business intelligence implementations are based on operational data.
that can be considered as more favorable than measurements or tests performed for
design purposes. Burden is a key design parameter in blasthole design; however, its
impact on fragmentation is currently considered in a deterministic manner. Geospatial
data sourced from drill navigation systems can be utilized in real-time.

4.5.1. Geospatial data manipulation of drill navigation system data

The use of geospatial data in information systems is not a new technology having an
application history of more than twenty years. The complexity of spatial data was
originally limited by highly specialized user profiles, which restricted its potential to be
implemented in data integration. Information becomes more valuable when it is
integrated with other data sources and can be accessed via easy-to-use environments. Database management systems enhanced their solutions by
developing spatial data tools; however, the skills required to effectively utilize them is
still limited. A solid understanding of basic spatial data concepts is very important
prior to implementing any kind of data integration (Alastair, 2009).

The mine-to-mill process involves different operational stages and their related data,
a considerable amount of which can be classified as geospatial. Starting with the
drilling process, data provided by drill monitoring systems provides drillhole locations.
Points that have no area or length represent exact locations. Drillhole locations are a
good example of points that could be defined on the database layer. This data is
utilized in calculations related to blasting and fragmentation. However, empirical
approaches lose the granularity of data by relying on a limited amount of data.
Burden and spacing calculations can be performed on real-time data when a data
warehouse is available. There are also other spatial data types such as lines and
polygons that can be used to augment mining related data.

Drill navigation systems provide such measures as penetration rate, torque, and
revolutions per minute (rpm) during operation together with precise GPS locations for
each hole. This study utilized geospatial indexing and available functions in MS SQL
Server to handle drill and blast related data. Drillhole coordinates were converted into
a spatial data type, similar to a string, and stored in the same fact table. This string
joined X and Y coordinates with a spatial reference identifier (SRID) giving the fact
table a geospatial meaning. Once the geospatial column was stored, visualizing the
drillholes was possible using the RDBMS, as shown in Figure 4.8. In addition, other applications could also be used to pinpoint the drillholes on a map layer.

![Geospatial visualization of drillhole data in RDBMS](image)

Figure 4.8 Geospatial visualization of drillhole data in RDBMS

In order to improve the performance of geospatial queries, indexing was performed. A stored procedure was developed to read the X and Y coordinates of a drillhole, convert them into a geospatial string, and then apply a nearest neighborhood algorithm. The stored procedure parameters are the maximum distance and the number of drillholes to be searched by the algorithm. The nearest drillholes and their corresponding data in the fact table such as explosive amount and F80, were aggregated so that this information could be joined back to the original fact table. In this manner, adding measures representing close drillholes enhanced the data for every record in the fact table. Cursor logic was implemented to run the stored procedure against the whole table, row by row. The performance of using geospatial data, index, and functions proved that the built-in properties of MS SQL Server are reasonable alternatives to developing complex queries.
4.6. Deployment of the data mining model to the data warehouse

Data mining models were trained to predict the fragmentation based on drilling and blasting related data. The final stage of the suggested framework is deployment of data mining models into the same environment in which operational data is stored. Data mining models are structured to enable multiple rules and variable definitions to be used for representation. Predictive model markup language (PMML) is a standard that is supported by software commonly used for data mining and statistical models (DMG, 2015). XML is used to represent the data mining models; a schema defines its structure. A partial PMML sample of a decision tree model trained on the integrated data can be seen in Table 4.6.

Table 4.6 PMML sample of a decision tree model

```
<Header>
<DataDictionary numberOfFields="14">
<DataField name="Fragmentation" optype="categorical" dataType="string">
  <Value value="Coarse"/>
  <Value value="Fine"/>
</DataField>
...
<TreeModel modelName="RPart_Model" functionName="classification">
  <algorithmName="rpart" splitCharacteristic="binarySplit"
    missingValueStrategy="defaultChild">
    <MiningSchema>
      <MiningField name="Fragmentation" usageType="predicted"/>
      <MiningField name="Explosive.Type" usageType="active"/>
    </MiningSchema>
    <Node id="29" score="Fine" recordCount="81" defaultChild="58">
      <SimplePredicate field="DipDist" operator="lessThan" value="3.837325"/>  
      <ScoreDistribution value="Coarse" recordCount="28" 
        confidence="0.345679012345679"/>  
      <ScoreDistribution value="Fine" recordCount="53" 
        confidence="0.654320987654321"/>
    </Node>
  </TreeModel>
</PMML>
```

As seen in Table 4.6, the PMML schema defines the input variables, model parameters, and details, such as nodes for the decision tree model, in a structured way. This enables models trained in a software package to be exported to other data mining environments or even an RDBMS. The integrated data warehouse used for mine-to-mill optimization has the potential to be enhanced by implementing data mining models developed in other software. Modern mines could benefit from data
mining models in their data warehouses by re-training and re-evaluating them using real-time data. PMML can be transformed into SQL, as shown in Table 4.7.

Table 4.7 SQL script generated from PMML of the decision tree model

```
SELECT b.*
    ,CASE
      WHEN ENDNODE_NUM = 4 THEN 0.705089820359281
    ... END as TARGETSCORE_Coarse_CONFIDENCE
FROM [table name] a) b
```
5. RESULTS OF DATA MINING MODELS FOR MINE-TO-MILL OPTIMIZATION

Data integrated from several systems was enhanced by geospatial queries in a real-time data warehouse. Data was prepared for the data mining application by implementing various data cleaning and data enhancement techniques. Once data was cleaned by filtering out missing and erroneous records from all integrated data sources, the scaled down data frame was enhanced by using geospatial queries. Amount and quality of potential input variables were improved prior to data mining application.

The results of the data mining implementation are divided into two case studies that are trained for three different data mining models with two different threshold values. Decision tree, random forest, and adaptive boosting algorithm models were trained on the same data set that was classified as fine and coarse by their average \( F_{80} \) values associated to blastholes. Two different thresholds were used to classify the target variable of fragmentation. As a result of detailed investigation of all available data sources and communication with managers on site, the threshold value of 7" was used as it was a goal set in the balanced scorecard used to evaluate the drill and blast manager performance. This value is aimed to be reached after blasting material and is based on the size distribution that the drill and blast assessment system provides by processing photos taken at the shovel location.

Case-A is solely based on data collected from automated systems such as the drill navigation system, the laboratory information management system, fleet management system, and others. The same data set was used to integrate manually collected data related to drilling and blasting for Case-B. This manually collected data set covers average burden and spacing for a shot, powder factor, and other variables. Integration of the manually collected data was performed by associating variables defined for a shot to all the blastholes that were covered in this group. This resulted in having additional variables that could be used in the training phase of data mining models. However, the lack of variance in the data representing burden, spacing, rock type, and rock density resulted in underutilization of these variables throughout building and validating data mining models. This issue is covered in more detail in the next chapter where solutions are also suggested.
5.1. Case Study-A

This case study was built on data that was generated by and obtained from systems related to drilling and blasting. The data warehouse provided by the open pit copper mine was cleaned and enhanced to prepare a data frame for training data mining models. All variables that were chosen as potential input to data mining models were investigated for their characteristics and distributions representing these variables were reviewed by histograms. Figure 5.1 depicts the distribution of one of the model input parameters as an example where average explosive amount loaded into blastholes was considered. All parameters were investigated for outliers and preliminary data filtering to enhance data prior to the data mining implementation.

![Figure 5.1 Distribution of average explosive amount](image)

Average explosive amount loaded in blastholes is an important variable by its impact on defining explosive energy and powder factor related to each blasthole. A contractor at this mine performs explosives' loading and it can be seen that there is a considerable variation in amount loaded. Blastholes that have missing values and outliers for average explosive amount were removed from the data set prior to modeling. Fragmentation was set as the target variable and determined by a threshold value. Two different scenarios, in which the target fragmentation was 5 inches and 7 inches were used to categorize the average F<sub>80</sub> measure as fine and coarse. As shown in Figure 5.2, the fine category dominates the data set for a threshold value of 7 inches.
The drilling and blasting department scorecard was used to determine the threshold value of 7 in as it was one of the key measures that the manager was evaluated by. The goal defined as target fragmentation at shovel is a proactive measure that can be used to assess the performance of supervisors. Blastholes that have data associated with an average fragmentation of less than 7 inches material is more than material greater than the threshold. In other words, amount of fine material is approximately 1.7 times more than coarse material. Using the target variable of average $F_{80}$ value associated to blastholes in a categorical way enabled the utilization of more data mining models in Rattle. Available data mining models were first trained for predicting the target variable classified by a threshold of 7 in. and then 5 in. This supports the idea of having different case studies, which are mainly different from each other by the amount of fine and coarse material. Re-classifying the data frame by using different threshold values provided a comparison basis for the performance of the three data mining models trained on the same data sets in this study.

Figure 5.2 Distribution of fragmentation data for 7 in.
The first data mining model trained for the case where automated data was used with a threshold of 7 in. was the decision tree model. Figure 5.3 illustrates the decision tree that used average explosive amount, minimum distance, diameter, and hole depth to create primary splits in the data. Decision trees can easily be interpreted to show which variables are used to create classification rules.

Tree construction starts with a criterion, associated with the minimum distance, that could be considered as an alternative to the concept of burden; however, the criterion is not based on an empirical approach. The decision tree performs a better split in the second level, having a lower cross-error, where average explosive amount is used to classify data as fine and coarse. Another classification is then based on the variable dipped length/distance that indicates the role of blasthole depth. The last branch is
formed by using average explosive amount but is a comparably less important variable in this decision tree. The overall error rate for the decision tree was 33% when the training data was used for evaluation.

A random forest was built for the same data to compare the overall error rate with decision trees. As multiple trees are built without pruning, the visualization of random trees is not as practical as that of a single decision tree. Therefore, importance of variable, as seen in Figure 5.4, represents how frequently certain inputs were utilized and clarifies the prediction process of random forests.

![Figure 5.4 Variable importance for random forest model (7 in.) for Case-A](image_url)

Variable importance is based on how frequently a variable was used in constructing decision trees within a random forest. Similar to the naïve approach of building a single decision tree, random forests indicate that average explosive amount and
minimum distance to a neighboring blasthole are two of the most important variables used. Diameter and depth of blastholes, combined with the distances that we calculated by using geospatial queries, have lower scores indicating their less frequent usage in building the random forest. Average water depth was marked as the least important variable for the random forest model in this case. This could be related to the water depth showing low variance for recorded for blastholes that are classified as fine. The overall error rate is reduced to 31%, and the prediction of coarse particles, larger than 7 in., is considered to be the major cause of error.

Adaptive boosting is another modeling method that utilizes the concept of building smaller trees and promoting variables in a continuous manner. Weights are assigned in a way that variables that were not used in building a tree are prioritized. As multiple decision trees were built until no observation was left to be boosted, plotting this model is also not practical. As seen in Figure 5.5, the most important variables in this model are related to blast pattern geometry, which were enhanced by the distances calculated by the nearest neighborhood algorithm.

![Figure 5.5 Variable importance for adaptive boosting model (7 in.) for Case-A](image.png)
In contrast to the random forest model, the average amount of explosive used in the adaptive boosting algorithm was not classified as the most important variable that improved the performance of prediction. Stemming and diameter were variables that were identified as having the most impact on training the model and predicting fragmentation. Average water depth was again not marked as being effective on increasing the accuracy of the adaptive boosting algorithm. Variables that were enhanced by the distances calculated via using geospatial queries had higher utilization when compared to variables that define explosives loaded to each blasthole. The overall error was reduced to 29% with this model and indicated that the adaptive boosting algorithm performed better than the decision tree and the random forest in this case. Similar to random forests, the accuracy of this model was challenged by predicting coarse particles. Therefore, raw data was re-classified with a threshold value of 5 inches for another case, which resulted in the distribution as seen in Figure 5.6.

![Figure 5.6 Distribution of fragmentation data for 5 in.](image)
Data is a major actor that affects data mining models during the training and testing stages. The same data mining models were trained and evaluated for the scenario in which target fragmentation size at the shovel is 5 in. Amount of data classified as fine material reduced when compared to the case where the threshold was used as 7 in. This case study has a dominating amount of data classified as coarse; approximately two times the amount of fine data. Using these different threshold values aided to generate data sets with different distributions of fragmentation aiming to improve the accuracy in prediction.

Similar to the case where a threshold value of 7 in. was used to classify fragmentation related data; same three data mining models were trained and evaluated. The decision tree model, shown in Figure 5.7, has a lower error rate when compared to the case in which threshold of fragmentation was 7 inches.

Figure 5.7 Decision tree model (5 in.) for Case-A
In contrast to the decision tree built on a fragmentation threshold of 7 in., average water depth was determined to be an important variable in the model built for 5 in. This indicates the difference of values representing average water depth associated to blastholes that are related to coarse material having higher variance. Using a different threshold value changed the proportions of data associated to fine and coarse material. Average explosive amount and diameter of blastholes generated a secondary split of data. The overall error rate was 27%, which indicates that classifying fragmentation data with 5 in. improved the decision tree performance. Average explosive amount and diameter are the other major variables that were utilized to predict fragmentation in the model.

The random forest model was again trained for this case to observe the impact of data classified by its average F80 value with a threshold of 5 in. As illustrated in Figure 5.8 the variable importance plot was used to evaluate which parameters were used more frequently to build the random forest.

![Figure 5.8 Variable importance for random forest (5 in.) for Case-A](image)

Figure 5.8 Variable importance for random forest (5 in.) for Case-A
Similar to the case where the variable representing fragmentation was divided into two different groups by a threshold of 7 in., average explosive amount was again the most highly utilized variable in building the random forest model. As smaller decision trees are built to construct the random forest, variables that split data on first and second levels in the decision tree show higher utilization. Average water depth and stemming were inputs that the model utilized frequently during the training stage. Diameter and depth of blastholes had a higher score than average penetration rate and the minimum distance between blastholes. This points out the fact that data enhancement might have a positive impact on prediction success. The error rate of predicting fragmentation by a conditional random forest was 25%. This was the lowest error rate reached by using three different data mining models on two different data frame scenarios. The adaptive boosting algorithm was also trained for the same data frame; the important variables are shown in Figure 5.9.

![Variable importance for adaptive boosting model (5 in.) for Case-A](image)

Figure 5.9 Variable importance for adaptive boosting model (5 in.) for Case-A
Unlike the scenario in which input data was classified for 7 in., the adaptive boosting algorithm had a higher error rate of 29%. Since the major difference between the two scenarios is the distribution of fine and coarse data, the distribution of errors also showed variance. Reclassifying data improved the success rate in predicting fragmentation for decision trees and random forests. Drilled depth of blastholes was the most important variable used in the adaptive boosting algorithm. Stemming and diameter also classified as major variables, which supported the conclusion that blast design parameters enhanced by distance calculations had more impact than average explosives amount and average penetration rate. Prediction of actual fine fragments as coarse fragments was one of the main sources of error for the adaptive boosting algorithm. A conclusive comparison of the data mining models built with 7 in. data is shown on the ROC curve of Figure 5.10.

![ROC curve of models evaluated for 7 in. for Case-A](image)

Figure 5.10 ROC curve of models evaluated for 7 in. for Case-A
The area under the ROC curve is utilized to evaluate model performance. The area under the curve for the decision tree model was 0.55, lower than that of the random forest and adaptive boosting algorithms. A single decision tree was found to be insufficient to predict fragmentation, whereas more sophisticated models had an area of approximately 0.64 under the ROC curve. Similarly, in the 5 in. case, the ROC curve indicated that a difference in prediction performance among the three models, as illustrated in Figure 5.11.

![Figure 5.11 ROC curve of models evaluated for 5 in. for Case-A](image)

Similar to the case in which data with a threshold of 7 in. was used, the decision tree model underperformed in predicting fragmentation. The random forest and adaptive boosting algorithms had higher areas under the ROC curve, indicating higher true
positive rates. This reveals the potential of more sophisticated models to support decision-making in the mining industry. The amount and quality of data is key to improving prediction performance. This conclusion is summarized by Table 5.1 and Table 5.2. It reinforces the point that technology that generates, captures, and integrates operational data is crucial for modern mines throughout mine-to-mill implementations and data mining studies.

Table 5.1. Confusion Matrix for 5 in. scenario

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coarse</td>
<td>Fine</td>
</tr>
<tr>
<td>Decision Tree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ov. Error: 27%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarse</td>
<td>128</td>
<td>11</td>
</tr>
<tr>
<td>Fine</td>
<td>43</td>
<td>15</td>
</tr>
<tr>
<td>Random Forest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ov. Error: 25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarse</td>
<td>130</td>
<td>9</td>
</tr>
<tr>
<td>Fine</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Adap. Boosting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ov. Error: 29%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coarse</td>
<td>118</td>
<td>21</td>
</tr>
<tr>
<td>Fine</td>
<td>37</td>
<td>21</td>
</tr>
</tbody>
</table>

The three different data mining algorithms were applied to fragmentation data categorized by a threshold of 5 in. The decision tree model had an overall error of 27%, mainly caused by predicting actual fine data as coarse. The adaptive boosting algorithm had a higher error rate and underperformed in the prediction of coarse data. The random forest model had the best prediction performance, with an error of 25%, producing better results for coarse data than other models. A possible cause of having a better performance in predicting data classified as coarse might be the proportion that changed in favor of coarse material when a threshold of 5 in. was used. The performance of the same models trained on data categorized with a 7 in. threshold is summarized in Table 5.2.
Table 5.2 Confusion Matrix for 7 in. scenario

<table>
<thead>
<tr>
<th>Threshold:</th>
<th>7 in.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>Ov. Error: 33%</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Forest</td>
<td>Ov. Error: 31%</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Adap. Boosting</td>
<td>Ov. Error: 29%</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The adaptive boosting algorithm, which incrementally prioritizes weights, had a comparably lower error rate. Data classified as fine was mainly predicted true, whereas errors were seen in classifying coarse data. The random forests had a lower error rate than that of decision trees. This result validates the concept of building multiple trees instead of a single structure. The variance in the error rate of selected data mining models indicates that a single model might not provide the required accuracy in all cases. Deployment of data mining models into the same environment where the data is collected and integrated becomes more crucial as having all rules associated to different data mining models enables a dynamic analysis capability of the data warehouse. This way, prediction does not have to rely on a single model and a comparison could be made each time new data is added to the system and used in training data mining models.

5.2. Case Study-B
The data set used for Case Study-A was enhanced and verified by a manually collected data source. Additional input parameters, such as powder factor and rock density, were utilized in the analysis. The manually collected data was added to the data warehouse as a comparably small database. Although amount of data was less than the data gathered by automatic systems and the level of granularity was different, it was preferred to integrate this small dataset. This way, impact of adding manually collected data could be seen. Another difference of this case was the way average penetration rate was utilized in the data frame. As this variable was sourced
from the drill navigation system, it was considered as a major and potentially important input to predict fragmentation. All data mining models that were trained and evaluated in Case-A relied on data that was collected automatically and average penetration rate was not stated to be an important variable for training any of the models. Therefore, three different classes of rock strength were created to convert the numeric value of average penetration rate into classes aiming to improve its potential to increase the accuracy of prediction. Rock type and rock density were cross-checked with the laboratory information management system before integrating them. The decision tree built for this enhanced data frame with a threshold of 5 in. is represented in Figure 5.12.
The decision tree built for the 5" case is primarily split by the hole depth that was collected manually, which indicated there is potential to fill data gaps in automatically collected data from systems such as the drill navigation system by utilizing manually collected data. However, the issue of QA/QC related to manual data collection has to be implemented prior to any integration or analysis in future studies. Average water depth and diameter combined with the distances between neighboring blastholes are other variables that classified fragmentation data successfully. Hole depth that was derived from integrating the drill navigation system and explosives loading reports also aided classification on a lower level compared to manually collected hole depth. Combining these measures might result in a better performance for the decision tree. Rock strength that was derived from the average penetration rate was not efficient in splitting the tree and indicated that using more data sourced from the drill navigation system might be required. More sophisticated models than the decision tree were trained on the same data set and variable importance plot for the adaptive boosting model is shown in Figure 5.13.

![Variable importance for adaptive boosting model (5 in.) for Case-B](image)

Figure 5.13 Variable importance for adaptive boosting model (5 in.) for Case-B
The manually collected data about drilling and blasting operations on site had an impact on the variable importance. Hole diameter, stemming, and hole depth were transformed to have more sophisticated meanings by introducing the distances found via geospatial queries. These variables were still marked as important when powder factor and rock density were introduced in the model. This indicates that there might be additional variables that could benefit prediction. However, the amount and granularity of the manually collected data was limited when compared to the integrated data layer. Average burden and spacing that was recorded for a whole shot had the least frequency of usage while training the model. This supports the idea of generating new measures to associate them on a single blasthole level by changing the level of granularity. Values that are recorded as averages have low variance and were seen as important variables for the more sophisticated models.

Figure 5.14 provides the variable importance plot of the random forest model.
Hole depth, which was manually recorded, had the strongest effect on increasing the accuracy of the random forest model. Stemming, hole diameter, and minimum distance between neighboring blastholes were utilized more frequently than other input parameters. This plot indicates that blasthole design parameters might have a greater impact on fragmentation compared to explosive properties and the surrounding rock. Powder factor which defines the amount of explosive required to break a unit measure of rock had a comparably lower utilization by the random forest model during its training phase. The limited amount of data recorded to represent a shot of blastholes might be used to generate new measure by combining it distances calculated between neighboring blastholes. Change in explosive density or powder factor over a certain distance could be defined to emphasize the impact of explosives on fragmentation. All data mining models were compared by their true positive and false positive rates. The ROC curve of data mining models trained on data with a 5 in. threshold is shown in Figure 5.15.

Figure 5.15 ROC curve of models evaluated for 5 in. for Case-B
The random forest model performed better than the adaptive boosting algorithm and the decision tree based on the areas under curves representing prediction rates of all three data mining models. Decision tree models, considered to be more basic approaches, underperformed in representing the complex concept of mine-to-mill. True positive and false positive rates are affected by the way the target variable is classified. A threshold of 5 in. generates more data categorized as coarse; therefore, prediction of the fine class becomes prone to higher error rates. On the other hand, a threshold of 7 in. transforms the same data set to have a more crowded fine class.

The same data mining models were trained and evaluated by classifying data that was based on integration automatic systems and manually collected records by a threshold of 7 in. The decision tree model trained for this case is seen in Figure 5.16.

![Decision tree model (7 in.) for Case-B](image)

Figure 5.16 Decision tree model (7 in.) for Case-B
The same data set classified with a different threshold was used to build a decision tree that primarily depended on the measures that were enhanced by the geospatial queries. Minimum distance between holes, depth, diameter and stemming values over the distance between neighboring blastholes were used to classify fragmentation data. The main reason of these variables being used to create successful splits is the amount of unique values they have. Fragmentation models that utilize blasthole design parameters are commonly improved by modifying how the surrounding rock is defined whereas variables, such as depth and stemming are operational parameters. Geospatial queries and the nearest neighborhood algorithm used in this study enabled to give blasthole design parameters a more dynamic characteristic. This is of major importance as the data collected to represent the blasthole patterns will not show any variance if the mine prefers to use fixed plans for ore and overburden blasting. In addition, operations that have limited amount of technology and systems that collect data rely on manually collected records which could be enhanced by using them together with other variables. The random forest model was trained for same data and Figure 5.17 illustrates that variable in this case.

![Figure 5.17 Variable importance for random forest model (7 in.) for Case-B](image)
Similar to the model trained on data classified by 5 in., distance related blasthole design parameters had a higher potential to increase the accuracy of the random forest model. Manually collected data, such as powder factor and average burden, had a comparably lower utilization in the model; this might be caused by the low variance of these variables. Data mining models tend to utilize data with many unique values compared to variables that are limited, such as rock type, rock density, or average spacing in this case. As the scope of this study was set to focus only on the material that is sent to the concentrator, rock types were filtered down to low-grade sulfide ores. This also resulted in downgrading the impact of rock density that might be combined with other variables to increase its utilization in training data mining models. Figure 5.18 provides an evaluation of variable importance evaluation for the adaptive boosting algorithm trained on the same data set.

Figure 5.18 Variable importance for adaptive boosting model (7 in.) for Case-B
The importance of distance related variables was determined to be higher than that of manually collected blasthole design parameters, such as average burden and spacing. The explosive energy provided per ton of material had a comparably stronger impact on the model when compared to other manually collected variables. For instance, rock density had a very low utilization during the training phase of the model, which might be caused by having a limited number of unique values. This could also be interpreted as blastholes being drilled in homogeneous rock formation. Another possible interpretation might be that manually collected data requires more precise QA/QC. A well-structured strategy to record operational data related to drilling and blasting will aid maintaining a higher level of quality for data. The data mining models trained to predict fragmentation in this case compared in Figure 5.19 that illustrates the ROC curve.

Figure 5.19 ROC curve of models evaluated for 7 in. for Case-B
All data mining models trained on the data set with a threshold of 7 in. had similar performance. A common issue in all models was observed in the prediction of data classified as coarse. The random forest model had a slightly lower error rate than the other two models. The prediction of data classified as fine was performed in the most accurate way by the random forest model. A general conclusion of random forest model to have the highest potential to represent the complexity of mine-to-mill proves the point that basic classification models, such as the decision tree might remain insufficient. However, both random forests and adaptive boosting models benefit from decision trees that are built without pruning. Building a full scale decision tree might be negatively affected by over-training the data and result in depending on the data frame in an excessive way. Building models by following a recursive structure provides the flexibility to adapt new data in a more efficient and accurate scheme.
6. CONCLUSION AND RECOMMENDATIONS

Mine-to-mill is a challenging problem, given its complex structure built in numerous operational stages. Each stage in the value chain has parameters that affect product quality. Blasting fragmentation is an area commonly studied for potential improvement. The cost related to blasting is considered low as compared to modifying downstream processes, such as crushing or grinding. Therefore, optimal blasting fragmentation is key to success in mine-to-mill studies. Keeping track of data and its quality of production is a challenging task that supervisors and managers have to fulfill on a daily basis. Drilling and blasting are crucial stages of production that cover various processes requiring continuous control to prevent delays or down events. Lack of communication between departments and information to support decision making are main deficiencies that limit mine-to-mill initiatives. Available technology in the mining industry provides tools to create an environment to integrate different data sources following continuous improvement strategies.

Investment in technology is an inevitable stage of gaining a corporate identity that is commonly driven by the implementation of ERP systems in mining companies. Integration of data is a topic that is included in the scope of ERP systems, mine planning softwares, or enterprise infrastructures. However, this requires an expertise in the field for which data is captured and is commonly led towards the perspective of the integrating platform. Modern mines need a detailed analysis of available data and plans about utilizing data once it is integrated. This study gave the basics to construct a data driven framework for mining engineering supported by real-time data of an integrated data warehouse. An integrated data warehouse was a fundamental component of this study, as real-time integrated data determined the success of the research. Different sources of live data were utilized to determine data relationships or patterns significant to the mine-to-mill process. One of the major advantages of implementing data warehousing methodology at mines is that a backup of clean data is created and could be used for historical analysis or as part of unexpected data loss due to failures in the infrastructure. Tools that are available in the data warehouse provide both analytics and reporting features in the same environment. These are the basis of utilizing data on a continuous basis and changing management perspective of daily operations.
The framework provides an end-to-end path to adopt in order to benefit from technology in modern mining operations. Basis of all BI implementations and data mining applications is the integrated layer of data. The mine already had the corporate data warehouse in place so the framework pointed out best practices in building this infrastructure. However, integrating data and implementing BI solutions into daily operation improved data utilization both via using OLAP cubes and balanced scorecards. Communication between departments improved, common goals shared for both mining and mineral processing influenced how operational decisions were made. The suggested framework provided a guideline to mining engineers for realizing the potential benefit of using data. Depending on data becomes common practice among all levels of operations once the BI implementations earn the trust of managers. Utilization of tools, such as OLAP cubes and balanced scorecards that were presented in this study as part of the framework, creates corrective action.

Visualizing data in balanced scorecards and providing a flexible environment to conduct analysis in OLAP cubes exposed issues in daily operation. Effectively using drill navigation data revealed differences between design and actual drill locations. This way, drill equipment accuracy was improved and kept under a pre-determined error rate used in scorecards. Similarly, the BI application based on the daily report provided by a contractor led to improvement in explosive loading. Amount and density of explosive loaded into each blasthole was tracked as a QA/QC measure. Accounting of explosive cost was supported by real data and discrepancies causing an unbalanced distribution of explosive energy on site were handled with the help of the integrated mine system defined in the framework. Role of people is crucial to drive corrective action and increase data quality. Mining engineers are highly valuable assets that should preferably spend time and effort in analyzing data and increasing productivity. Data collection is not a task that can be performed to efficiently together with daily tasks that mining engineers complete. Therefore, a reliable source of data as the data warehouse used in this study should be available to mining engineers that are trained for analyzing data. Expanding the vision of mining engineers throughout their education to get familiar with technology and data for continuous improvement in mines will aid the success in following the proposed framework.
Geospatial queries were used to enhance existing data related to drilling and blasting prior to training of data mining models. Storing mining related data in geospatial context and indexing this data by defining boundaries, introduced the potential of it to handle data sourced from drill navigation systems and fleet management systems. The execution time and reduced complexity of scripting geospatial queries was presented in this study. Location based data is generated by many different ways mainly using hardware equipped with GPS and that is capable of communicating via Wi-Fi or Bluetooth. Using data together with its location has advantages for the mining engineer that analyzes it. Mines operate on frequently changing layouts that emphasizes the importance of using precise location of equipment and people. The geospatial queries that were developed for the nearest neighborhood algorithm in this study used some of the in-built functions in SQL Server. It is also possible to create geometries that cover certain areas of operation and used for both production and safety purposes. There are already implementations of data that are used to track personnel and equipment, such as fleet management systems. However, handling this highly frequent data in a conventional way requires additional effort to optimize storage and performance of queries in the data warehouse. Data types and indexing conducted in this study indicated that using data with a geospatial meaning is a powerful option for mines equipped with technology.

Data mining proved itself as an advanced analysis tool that could be integrated with the data warehouse to dynamically use real-time data. The deployment of data mining models into the data warehouse closed the loop of data flow and enriched the infrastructure for further analysis. Volume, variety, and velocity of data generated in modern mines points to the fact that mining industry is progressing towards the big data concept. The dynamic analysis of data is important as time is valuable resource for all production stages of mining. Data mining models that are trained and evaluated on a continuous basis built up a core piece of the proposed framework in this study. Development of hardware and software to handle streaming data will aid mining operations to adapt new technology. Live data streams could be used to for analysis while they are collected in contrast to storage in intermediate tables. This exposes the importance of data quality for automatic systems and manually collected records. Data has to reliable both for analysis and reporting purposes that provide a basis for modern mine management.
The case studies used in the data mining implementation supports the requirement of better quality control for manually collected data. Although the level of granularity is different, manually collected data has the potential to fill in the missing pieces that are tracked by systems. This also indicates the importance of people that collect and utilize data. All levels of operational staff, such as managers, supervisors, and operators should be familiar with using the available data either by BI tools that can be developed or pivot tables in spreadsheets. Giving access to a reliable source of data leads to personnel adapt the data-to-action concept in the most efficient way. Performance assessment that is based on predetermined goals generates a perspective to evaluate operational problems on a more short-term basis and conduct root-cause analysis by drilling down into the available layer of integrated data. This can be considered as a two-folded benefit as people that are equipped with the right training to analyze data will also initiate corrective action to fix systems that capture data on site.

Amount of data generated in modern mines will continue to increase, as technology is a crucial piece of operation. Wearable technology and Internet of things are concepts that are not only related to sensors and gears. Data mining has the ability to handle a vast amount of data, either structured or non-structured. Data integration and analytics will be more commonly used in performance and safety metrics of workforce. The data mining application for mine-to-mill optimization in this study proved itself as an alternative to analyze data generated by new technology implementations. Although this study used a data warehouse of a copper mine that has a level of technology awareness and data infrastructure, operations that have lower data utilization and available technology could still benefit from the defined framework.

Mines that operate on a comparably lower scale are challenged by the amount of technology available on site. Although technology investment is a main part of this framework, potential of manually collected data indicates that utilization of data could benefit significantly from following a strategy to collect, clean, and enhance data. Data cleaning and enhancement presented in this study are guided by the mindset that is aware of the improvement potential of data. Being familiar with the technology that generates and captures data provides the fundamental background to handle discrepancies and missing records. Cleaning data is a task that requires expertise in
operations and input from people that are familiar with the process that is represented by data. Integration and enhancement of available data is based on a continuous improvement perspective can be initiated regardless of the data infrastructure in place. Having the outlook of using data will pioneer all related steps that have to be completed in order to build up an integrated layer of data. A detailed QA/QC process and pre-defined measures are essential for mines that benefit from a limited amount of technology.

The outcomes of this study provide a better understanding of mine-to-mill practices and enable future research in this field.

6.1. Future Research
This study builds up a framework for modern mines that aim to implement data-driven technology in their operations. Scope of the study was defined for a surface copper mine with a mine-to-mill initiative. Main limitation of implementing data warehousing and data mining at mine operations is the required infrastructure. The analysis capability of following this framework is proportional to the amount and quality of data collected. Topics planned to be studied in future research can be listed as follows:

- Importance of variables related to rock mechanics appeared weaker in data mining algorithms trained to estimate fragmentation. Lack of a geotechnical department and any previous consulting work conducted at the mine might have limited the representation of rock strength. Data collected by drill monitoring systems, such as penetration rate, torque, load on bit, is planned to be analyzed in detail to derive a rock strength parameter.

- Scope of the study was defined for mine-to-mill optimization whereas mine-to-leach is another concept that should be considered. The framework defined throughout the study has a generic approach to suit different cases. Operations that base their production on leaching have different target fragmentation, which can again be handled by integrated data. Data mining algorithms will be trained with different variables to estimate fragmentation or other key parameters that are crucial for mine-to-leach.
• The data gap between mining and mineral processing limited this study in incorporating process-based data. Energy consumption at the mill is a target variable that should be studied in detail. The impact of blasting quality decreases throughout size reduction stages, as the material gets crushed and grinded. Downstream processes, such as floatation should also be considered in data mining models for optimal fragmentation prediction. Material tracking needs to be performed in a more comprehensive way to track it continuously. Stockpile management is a complex stage of operation that is planned to be investigated both for mine-to-mill and grade control purposes.

• The role of people in an integrated mine system was discussed partially in the scope of this study. Engagement of continuous improvement in daily operations, performance metrics based on data utilization and data quality, impacts of business intelligence applications on operational cost are some topics that will be covered in other research studies. Benefits of following a data-to-action initiative will be supported by financial and technical analysis in detail.

• Conventional design of blasthole patterns is based on empirical equations. Although this study introduced the potential of geospatial queries and data sourced from drill monitoring systems, new measures could be derived and used for data mining purposes. Mine planning software has highly precise coordinates of the bench toe and crest in a vector format. Points on these strings could be used for a more realistic definition of burden where the drill monitoring system would be integrated with the mine planning software.

• Another research area is in enhancing the block model with the support of a data warehouse. This way, the block model used for short-term or long-term planning and optimization might get a more real-time role in operation. Although operational data would not be used for any kind of estimation in the block model, visualizing it together with mineral content will support scheduling and planning studies. Mining engineers that are responsible of planning and optimization will get insight about daily operations by having a block model improved by operational data. Block models used in a dynamic way could
benefit from historical data stored in the data warehouse to evaluate the success of their estimation by analyzing the impact on various processes.

- The error rate reached with the data mining models will be reduced by training the models with new measures and more data that will be available in the data warehouse within time. Association models will be trained to determine which parameters are relatively stronger related to the target variable and used in multiple regression analysis.

- A detailed cost analysis is planned to be conducted to compare blasting with comminution to come up with a decision support tool. Cost related to both processes will be investigated from a life-cycle perspective to cover all downstream processing costs. Similarly, a life cycle assessment study will be conducted to compare environmental impacts of different operational stages within the mine-to-mill concept according to their global warming potential, acidification values and other emission related impact categories.

- Additional innovative BI tools and data representations will be developed on production related data that will be integrated geospatially. On-the-fly calculations for blasthole design parameters, such as burden and spacing, are a primary focus. Near real-time feedback from the data warehouse might be used to improve drilling and explosive loading stages before ore is produced and handled by loading equipment. A heat map in SSRS will be used to indicate inefficiencies throughout the operation if up-to-date mine layouts are provided.
7. APPENDIX A: Impact of Different Type of Material on SAG Mill Performance and Feed Fragmentation
8. **APPENDIX B: Impact of Different Type of Material on Crusher Performance and SAG Feed Fragmentation**
9. APPENDIX C: Distribution of Different Type of Materials to Varying Dump Locations
10. APPENDIX D: Flowsheet of Crushing and Grinding Circuit
11. APPENDIX E: Selected SQL Queries

-- SQL script for decision tree model --

PMML2SQL generated SQL

Please read this before implementing SQL code:

* Search for '[table name]' in generated SQL code - and replace with name of score table.

* Make sure to validate SQL generated scores against scores from mining tool (on training data) before deploying SQL into production.

* The SQL code will not work correctly if the mining model uses variables that are altered or created within the data mining tool - unless same changes are made to the score data.

SELECT b.*
  , CASE
      WHEN ENDNODE_NUM = 4 THEN 0.705089820359281
      WHEN ENDNODE_NUM = 5 THEN 0.25
      WHEN ENDNODE_NUM = 12 THEN 0.738095238095238
      WHEN ENDNODE_NUM = 26 THEN 0.7
      WHEN ENDNODE_NUM = 27 THEN 0.285714285714286
      WHEN ENDNODE_NUM = 56 THEN 0.846153846153846
      WHEN ENDNODE_NUM = 57 THEN 0.363636363636364
      WHEN ENDNODE_NUM = 116 THEN 0.785714285714286
      WHEN ENDNODE_NUM = 117 THEN 0.320754716981132
      WHEN ENDNODE_NUM = 59 THEN 0
      WHEN ENDNODE_NUM = 15 THEN 0
      END as TARGETSCORE_Coarse_CONFIDENCE
  , CASE
      WHEN ENDNODE_NUM = 4 THEN 0.294910179640719
      WHEN ENDNODE_NUM = 5 THEN 0.75
WHEN ENDNODE_NUM = 12 THEN 0.261904761904762
WHEN ENDNODE_NUM = 26 THEN 0.3
WHEN ENDNODE_NUM = 27 THEN 0.714285714285714
WHEN ENDNODE_NUM = 56 THEN 0.153846153846154
WHEN ENDNODE_NUM = 57 THEN 0.636363636363636
WHEN ENDNODE_NUM = 116 THEN 0.214285714285714
WHEN ENDNODE_NUM = 117 THEN 0.679245283018868
WHEN ENDNODE_NUM = 59 THEN 1
WHEN ENDNODE_NUM = 15 THEN 1
END as TARGETSCORE_Fine_CONFIDENCE
FROM
(SELECT a.*, CASE
WHEN 1=1 AND Average.of.Explosive.Amount..SWE.<2136.833 AND DrillDist>=1.220178 THEN 4
WHEN 1=1 AND Average.of.Explosive.Amount..SWE.<2136.833 AND DrillDist<1.220178 THEN 5
WHEN 1=1 AND Average.of.Explosive.Amount..SWE.>=2136.833 AND Explosive.Type IN ('260-1') AND Average.of.Explosive.Amount..SWE.>=2187 THEN 12
WHEN 1=1 AND Average.of.Explosive.Amount..SWE.>=2136.833 AND Explosive.Type IN ('260-1') AND Average.of.Explosive.Amount..SWE.<2187 AND DiameterDist>=0.7780828 THEN 26
WHEN 1=1 AND Average.of.Explosive.Amount..SWE.>=2136.833 AND


WHEN 1=1 AND Average.of.Explosive.Amount..SWE.>=2136.833 AND Explosive.Type IN ('260-4','470-X') AND PitOpsPowderFactor>=1.479942 THEN 15

END AS ENDNODE_NUM

FROM [table name] a) b

-- PMML script for decision tree model --

<PMML version="4.2" xmlns="http://www.dmg.org/PMML-4_2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.dmg.org/PMML-4_2 http://www.dmg.org/v4-2/pmml-4-2.xsd">

  <Header copyright="Copyright (c) 2015 xavier" description="RPart Decision Tree Model">

    <Extension name="user" value="xavier" extender="Rattle/PMML"/>

    <Application name="Rattle/PMML" version="1.4"/>

    <Timestamp>2015-06-09 18:46:20</Timestamp>

  </Header>

  <DataDictionary numberOfFields="14">

    <DataField name="Fragmentation" optype="categorical" dataType="string">

      <Value value="Coarse"/>

      <Value value="Fine"/>

    </DataField>

    <DataField name="Explosive.Type" optype="categorical" dataType="string">

      <Value value="260-1"/>

  </DataDictionary>

  </PMML>
<Value value="260-4"/>
<Value value="462-X"/>
<Value value="470-X"/>
<Value value="None"/>
</DataField>
<DataField name="Average.of.Explosive.Amount..SWE." optype="continuous" dataType="double"/>
<DataField name="DiameterDist" optype="continuous" dataType="double"/>
<DataField name="DipDist" optype="continuous" dataType="double"/>
<DataField name="DrillDist" optype="continuous" dataType="double"/>
<DataField name="StemDist" optype="continuous" dataType="double"/>
<DataField name="PitOpsRockType" optype="categorical" dataType="string">
  <Value value="Diabase"/>
  <Value value="Schist"/>
</DataField>
<DataField name="PitOpsRockDensity" optype="continuous" dataType="double"/>
<DataField name="PitOpsAvgBurden" optype="continuous" dataType="double"/>
<DataField name="PitOpsAvgSpacing" optype="continuous" dataType="double"/>
<DataField name="PitOpsPowderFactor" optype="continuous" dataType="double"/>
<DataField name="PitOpsPowderKcal.Ton" optype="continuous" dataType="double"/>
<DataField name="Rock.Strength" optype="categorical" dataType="string">
  <Value value="Medium"/>
  <Value value="Strong"/>
  <Value value="Weak"/>
</DataField>
</DataDictionary>
<TreeModel modelName="RPart_Model" functionName="classification"
<MiningSchema>
  <MiningField name="Fragmentation" usageType="predicted"/>
  <MiningField name="Explosive.Type" usageType="active"/>
  <MiningField name="Average.of.Explosive.Amount..SWE." usageType="active"/>
  <MiningField name="DiameterDist" usageType="active"/>
  <MiningField name="DipDist" usageType="active"/>
  <MiningField name="DrillDist" usageType="active"/>
  <MiningField name="StemDist" usageType="active"/>
  <MiningField name="PitOpsRockType" usageType="active"/>
  <MiningField name="PitOpsRockDensity" usageType="active"/>
  <MiningField name="PitOpsAvgBurden" usageType="active"/>
  <MiningField name="PitOpsAvgSpacing" usageType="active"/>
  <MiningField name="PitOpsPowderFactor" usageType="active"/>
  <MiningField name="PitOpsPowderKcal.Ton" usageType="active"/>
  <MiningField name="Rock.Strength" usageType="active"/>
</MiningSchema>

<Output>
  <OutputField name="Predicted_Fragmentation" optype="categorical" dataType="string" feature="predictedValue"/>
  <OutputField name="Probability_Coarse" optype="continuous" dataType="double" feature="probability" value="Coarse"/>
  <OutputField name="Probability_Fine" optype="continuous" dataType="double" feature="probability" value="Fine"/>
</Output>

<Node id="1" score="Coarse" recordCount="858" defaultChild="2">
  <True/>
</Node>
<Node id="6" score="Coarse" recordCount="66" defaultChild="12">
  <SimpleSetPredicate field="Explosive.Type" booleanOperator="isIn">
    <Array n="1" type="string">"260-1"</Array>
  </SimpleSetPredicate>
  <ScoreDistribution value="Coarse" recordCount="42" confidence="0.636363636363636"/>
  <ScoreDistribution value="Fine" recordCount="24" confidence="0.363636363636364"/>
</Node>

<Node id="12" score="Coarse" recordCount="42">
  <SimplePredicate field="Average.of.Explosive.Amount..SWE." operator="greaterOrEqual" value="2187"/>
  <ScoreDistribution value="Coarse" recordCount="31" confidence="0.738095238095238"/>
  <ScoreDistribution value="Fine" recordCount="11" confidence="0.261904761904762"/>
</Node>

<Node id="13" score="Fine" recordCount="24" defaultChild="26">
  <SimplePredicate field="Average.of.Explosive.Amount..SWE." operator="lessThan" value="2187"/>
  <ScoreDistribution value="Coarse" recordCount="11" confidence="0.458333333333333"/>
  <ScoreDistribution value="Fine" recordCount="13" confidence="0.541666666666667"/>
</Node>

<Node id="26" score="Coarse" recordCount="10">
  <SimplePredicate field="DiameterDist" operator="greaterOrEqual" value="0.7780828"/>
  <ScoreDistribution value="Coarse" recordCount="7" confidence="0.7"/>
  <ScoreDistribution value="Fine" recordCount="3" confidence="0.3"/>
</Node>
<Node id="27" score="Fine" recordCount="14">
    <SimplePredicate field="DiameterDist" operator="lessThan" value="0.7780828"/>
    <ScoreDistribution value="Coarse" recordCount="4" confidence="0.285714285714286"/>
    <ScoreDistribution value="Fine" recordCount="10" confidence="0.714285714285714"/>
</Node>

<Node id="7" score="Fine" recordCount="112" defaultChild="14">
    <SimpleSetPredicate field="Explosive.Type" booleanOperator="isIn">
        <Array n="2" type="string">"260-4" "470-X"</Array>
    </SimpleSetPredicate>
    <ScoreDistribution value="Coarse" recordCount="43" confidence="0.383928571428571"/>
    <ScoreDistribution value="Fine" recordCount="69" confidence="0.616071428571429"/>
</Node>

<Node id="14" score="Fine" recordCount="105" defaultChild="28">
    <SimplePredicate field="PitOpsPowderFactor" operator="lessThan" value="1.479942"/>
    <ScoreDistribution value="Coarse" recordCount="43" confidence="0.40952380952381"/>
    <ScoreDistribution value="Fine" recordCount="62" confidence="0.59047619047619"/>
</Node>

<Node id="28" score="Coarse" recordCount="24" defaultChild="56">
    <SimplePredicate field="DipDist" operator="greaterOrEqual" value="3.837325"/>
    <ScoreDistribution value="Coarse" recordCount="15" confidence="0.625"/>
    <ScoreDistribution value="Fine" recordCount="9" confidence="0.375"/>
</Node>

<Node id="56" score="Coarse" recordCount="13"/>
<SimplePredicate field="DipDist" operator="lessThan" value="4.420916"/>

<ScoreDistribution value="Coarse" recordCount="11"
confidence="0.846153846153846"/>

<ScoreDistribution value="Fine" recordCount="2"
confidence="0.153846153846154"/>

</Node>

<Node id="57" score="Fine" recordCount="11">
  <SimplePredicate field="DipDist" operator="greaterOrEqual"
value="4.420916"/>

  <ScoreDistribution value="Coarse" recordCount="4"
confidence="0.363636363636364"/>

  <ScoreDistribution value="Fine" recordCount="7"
confidence="0.636363636363636"/>

</Node>

</Node>

<Node id="29" score="Fine" recordCount="81" defaultChild="58">
  <SimplePredicate field="DipDist" operator="lessThan" value="3.837325"/>

  <ScoreDistribution value="Coarse" recordCount="28"
confidence="0.345679012345679"/>

  <ScoreDistribution value="Fine" recordCount="53"
confidence="0.654320987654321"/>

  <Node id="58" score="Fine" recordCount="67" defaultChild="116">
    <SimplePredicate field="DipDist" operator="lessThan" value="3.182669"/>

    <ScoreDistribution value="Coarse" recordCount="28"
confidence="0.417910447761194"/>

    <ScoreDistribution value="Fine" recordCount="39"
confidence="0.582089552238806"/>

    <Node id="116" score="Coarse" recordCount="14">
      <SimpleSetPredicate field="Rock.Strength" booleanOperator="isIn">
        <Array n="1" type="string">"Weak"</Array>
      </SimpleSetPredicate>
    </Node>
  </Node>
</Node>
<SimpleSetPredicate>
  <ScoreDistribution value="Coarse" recordCount="11" confidence="0.785714285714286"/>
  <ScoreDistribution value="Fine" recordCount="3" confidence="0.214285714285714"/>
</SimpleSetPredicate>

<Node id="117" score="Fine" recordCount="53">
  <SimpleSetPredicate field="Rock.Strength" booleanOperator="isIn">
    <Array n="2" type="string">"Medium" "Strong"</Array>
  </SimpleSetPredicate>
  <ScoreDistribution value="Coarse" recordCount="17" confidence="0.320754716981132"/>
  <ScoreDistribution value="Fine" recordCount="36" confidence="0.679245283018868"/>
</Node>

<Node id="59" score="Fine" recordCount="14">
  <SimplePredicate field="DipDist" operator="greaterOrEqual" value="3.182669"/>
  <ScoreDistribution value="Coarse" recordCount="0" confidence="0"/>
  <ScoreDistribution value="Fine" recordCount="14" confidence="1"/>
</Node>

<Node id="15" score="Fine" recordCount="7">
  <SimplePredicate field="PitOpsPowderFactor" operator="greaterOrEqual" value="1.479942"/>
  <ScoreDistribution value="Coarse" recordCount="0" confidence="0"/>
  <ScoreDistribution value="Fine" recordCount="7" confidence="1"/>
</Node>
-- SQL script to create a fragmentation table and insert data –

USE [FRAG_geospatial]
GO

/****** Object: Table [dbo].[frag_cube_BRW_fltr] ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[frag_cube_BRW_fltr](
    [ID] [float] NOT NULL,
    [PK Date] [nvarchar](255) NULL,
    [Swe Shot Number] [float] NULL,
    [Primary Explosive] [nvarchar](255) NULL,
    [Blast Hole Name] [nvarchar](255) NULL,
    [DR Nav Collar Pos_X] [float] NULL,
    [DR Nav Collar Pos_Y] [float] NULL,
    [Hole Diameter (SWE)] [float] NULL,
    [F80 (Split)] [float] NULL,
    [Dipped Length (SWE)] [float] NULL,
    [Drilled Depth (SWE)] [float] NULL,
    [Stemmed Height(SWE)] [float] NULL,
    [Water Depth(SWE)] [float] NULL,
    [Explosive Amount (SWE)] [float] NULL,
    [Avg Pen Rate (DR Nav)] [float] NULL,
    [GeomLocation] [geometry] NULL,
    PRIMARY KEY CLUSTERED
    (    [ID] ASC
    )WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
    ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY] TEXTIMAGE_ON [PRIMARY]
GO

INSERT INTO [dbo].[Drillholes] ([ID],[PK Date],[Swe Shot Number],[Primary Explosive],[Blast Hole Name],[DR Nav Collar Pos_X],[DR Nav Collar Pos_Y],[F80 (Split)],[Hole Diameter (SWE)],[Dipped Length (SWE)],[Drilled Depth (SWE)],[Stemmed
VALUES ('1', '2014-01-13', 77, '470-X', '160017910/2610', 3063.1, 651.25, 1.61371195310.635, 46, 57.9986877418, 0, 1800, 221,06),

('2', '2014-01-13', '77', '470-X', '160017910/2611', 3077.17, 663.44, 3.020012617, 10.625, 18, 0, 2100, 114.96),

('3', '2014-01-13', '77', '470-X', '160017912/2621', 3061.83, 625.01, 1.764504967, 10.625, 60, 54.3897688, 18, 0, 2590, 286.81),

--Data insertion is executed to populate the table from an external source, the above given data is a sample due to confidentiality

GO

ALTER TABLE [dbo].[Drillholes]
ADD [GeoLocation] GEOGRAPHY
GO

UPDATE [dbo].[Drillholes]
SET [GeoLocation] = geography::STPointFromText('POINT(' + CAST([DR Nav Collar Pos_X] AS VARCHAR(20)) + ' ' + CAST([DR Nav Collar Pos_Y] AS VARCHAR(20)) + ')', 4326)
GO

ALTER TABLE [dbo].[Drillholes] ADD ID_PK INT PRIMARY KEY ([ID])

CREATE SPATIAL INDEX SIndx_Drillholes_GeoLocation
ON dbo.[Drillholes](GeoLocation)

-- Stored Procedure for calculating nearest neighborhood –

USE [FRAG_geospatial]
GO
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO

--select * from sys.all_columns where object_id = OBJECT_ID('nearest_points2')
ALTER PROCEDURE [dbo].[find_N_nearestCubeBrwsdFltr] @X FLOAT, @Y FLOAT, @maxdist FLOAT, @N INT
AS
BEGIN
DECLARE @pt GEOMETRY;


-- create point
SET @pt = Geometry::STGeomFromText('POINT(' + CAST(@X AS Varchar(40)) + ' ' + CAST(@Y AS Varchar(40)) + ')',4326);

-- return dataset
insert into [dbo].[nearest_pointsCubeBrwsFltr2]

SELECT TOP (@N) a.[Rank],
   a.[ID],
   a.[PK Date],
   a.[Swe Shot Number],
   a.[Primary Explosive],
   a.[Blast Hole Name],
   a.[DR Nav Collar Pos_X],
   a.[DR Nav Collar Pos_Y],
   a.[Hole Diameter (SWE)],
   a.[F80 (Split)],
   a.[Dipped Length (SWE)],
   a.[Drilled Depth (SWE)],
   a.[Stemmed Height(SWE)],
   a.[Water Depth(SWE)],
   a.[Explosive Amount (SWE)],
   a.[Avg Pen Rate (DR Nav)],
   a.[GeomLocation],
   a.dist
FROM
   (SELECT ROW_NUMBER() OVER(ORDER BY [GeomLocation].STDistance(@pt)) AS [Rank],
    [ID],
    [PK Date],
    [Swe Shot Number],
    [Primary Explosive],
    [Blast Hole Name],
    [DR Nav Collar Pos_X],
    [DR Nav Collar Pos_Y],
    [Hole Diameter (SWE)],
    [F80 (Split)],
    [Dipped Length (SWE)],
    [Drilled Depth (SWE)],
    [Stemmed Height(SWE)],
    [Water Depth(SWE)],
    [Explosive Amount (SWE)],
    [Avg Pen Rate (DR Nav)],
    [GeomLocation],
    [GeomLocation].STDistance(@pt) AS dist
FROM [dbo].[frag_cube_BRW_fltr] WITH (INDEX([SpatialIndex-fragCubeBrwsFltr])))
WHERE [GeomLocation].STDistance(@pt) < @maxdist
) AS a
ORDER BY a.dist;

END;

-- Cursor logic for geospatial query --
Declare @X int
Declare @Y int
Declare cur CURSOR LOCAL for

Select
   [DrillNavCollarPosX],
   [DrillNavCollarPosY]
from [dbo].[frag_cube2]

open cur

fetch next from cur into @X, @Y

while @@FETCH_STATUS = 0 BEGIN

EXEC [dbo].[find_N_nearest2] @X, @Y,50,6

fetch next from cur into @X, @Y
END

-- Script to create geospatial index --

USE [FRAG_geospatial]
GO

SET ARITHABORT ON
SET CONCAT_NULL_YIELDS_NULL ON
SET QUOTED_IDENTIFIER ON
SET ANSI_NULLS ON
SET ANSI_PADDING ON
SET ANSI_WARNINGS ON
SET NUMERIC_ROUNDABORT OFF
GO

/****** Object:  Index [SpatialIndex-fragCubeBrwsFltr]    Script Date: 17/6/2015
12:17:20 ******/
CREATE SPATIAL INDEX [SpatialIndex-fragCubeBrwsFltr] ON [dbo].[frag_cube_BRW_fltr] (GeomLocation)
)
)USING GEOMETRY_GRID
WITH (BOUNDING_BOX = (2800, -880, 4100, 2000), GRIDS = (LEVEL_1 = MEDIUM, LEVEL_2 = MEDIUM, LEVEL_3 = MEDIUM, LEVEL_4 = MEDIUM), CELLS_PER_OBJECT = 16, PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
12. REFERENCES


Data mining group, [Online] Available at: [http://www.dmg.org/v4-2-1/GeneralStructure.html](http://www.dmg.org/v4-2-1/GeneralStructure.html)


131


