



UNIVERSIDADE FEDERAL DE OURO PRETO

CURSO DE GRADUAÇÃO EM ENGENHARIA DE MINAS

**A análise e o uso de *Big Data* nas atividades
mineiras.**

**ESTUDO DE CASO: PROCESSAMENTO
MINERAL DE UMA MINA DE OURO NA AUSTRÁLIA**

Gustavo Henrique de Souza Crepalde

Ouro Preto

2017

Gustavo Henrique de Souza Crepalde

**A análise e o uso de *Big Data* nas atividades
mineiras.**

**ESTUDO DE CASO: PROCESSAMENTO
MINERAL DE UMA MINA DE OURO NA AUSTRÁLIA**

Monografia de Graduação defendida perante a banca examinadora, como parte dos requisitos necessários à aprovação na disciplina Trabalho de Conclusão de Curso II (TCC II) do Curso de Graduação de Engenharia de Minas.

Orientador: Prof. Dr. Helton Cristiano Gomes

Ouro Preto
Escola de Minas da UFOP
2017



Universidade Federal de Ouro Preto
Escola de Minas
Departamento de Engenharia de Minas – DEMIN
Campus Universitário – Morro do Cruzeiro
35400-000 – Ouro Preto - MG
Telefone: (31) – 3559-1590 – E-mail: demin@em.ufop.br
CURSO DE ENGENHARIA DE MINAS

A análise e o uso de *Big Data* nas atividades mineiras.

**ESTUDO DE CASO: PROCESSAMENTO
MINERAL DE UMA MINA DE OURO NA AUSTRÁLIA**

Autor: Gustavo Henrique de Souza Crepalde



MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
Universidade Federal de Ouro Preto
Escola de Minas - Departamento de Engenharia de Minas

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Aos 03 dias do mês de agosto de 2017, às 15h30min, no auditório do Departamento de Engenharia de Minas da Escola de Minas - DEMIN/EM, foi realizada a defesa do Trabalho de Conclusão de Curso de Engenharia de Minas requisito da disciplina MIN-491 – Trabalho de Conclusão de Curso II, intitulado “*A ANÁLISE E USO DE BIG DATA NAS ATIVIDADES MINEIRAS*”, pelo aluno **Gustavo Henrique de Souza Crepalde**, sendo a comissão avaliadora formada por **Prof. Dr. Helton Cristiano Gomes (orientador)**, **Prof. M. Sc. Magno Silvério Campos** e **Prof. M. Sc. Cristiano Luís Turbino de França e Silva**.

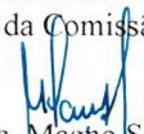
Após arguição sobre o trabalho, a comissão avaliadora deliberou por unanimidade pela **APROVAÇÃO**... do candidato, com a nota **9,0**... concedendo-lhe o prazo de 15 dias para incorporar no texto final da monografia as alterações determinadas/sugeridas pela banca.

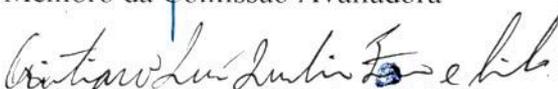
O aluno fará jus aos créditos e conceito de aprovação na disciplina MIN-491 – Trabalho de Conclusão de Curso II após a entrega dos exemplares definitivos (Cd e cópia impressa) da versão final da monografia defendida, conforme modelo do CEMIN-2009, no Colegiado do Curso de Engenharia de Minas – CEMIN.

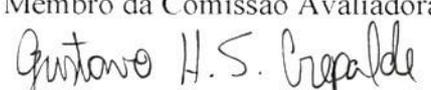
Para fins de registro, foi lavrada a presente ata que, depois de lida e aprovada é assinada pelos membros da comissão avaliadora e pelo discente.

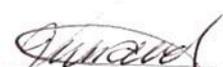
Ouro Preto, 03 de agosto de 2017.


Prof. Dr. Helton Cristiano Gomes
Presidente da Comissão Avaliadora e Professor Orientador


Prof. M. Sc. Magno Silvério Campos
Membro da Comissão Avaliadora


Prof. M. Sc. Cristiano Luís Turbino de França e Silva
Membro da Comissão Avaliadora


Gustavo Henrique de Souza Crepalde


Prof. Ms.C. José Fernando Miranda
Professor responsável pela Disciplina Min 491 – Trabalho de Conclusão de Curso

RESUMO

O sucesso de um empreendimento minerário está relacionado principalmente à produtividade da mina que consiste na relação de quanto é gasto com os trabalhadores, equipamentos e materiais e o quanto de minério é produzido por estes. Para que haja uma produtividade satisfatória é o crucial que o teor do minério trabalhado esteja dentro as especificações do mercado ou cliente. Este teor, muitas vezes não atende as necessidades do comprador por apresentar baixos índices do mineral minério e deve ser misturado a outro minério com alto teor em um processo chamado de *blending* ou mistura. Recentemente, a indústria mineral tem confrontado alguns desafios na manutenção de uma alta produtividade, pois com a escassez das reservas com altos teores de minérios, as empresas mineradoras têm trabalhado com teores cada vez mais baixos. Para manter uma produtividade satisfatória, bem como atender os anseios do mercado, as mineradoras estão investindo em novas tecnologias para otimizar suas atividades dentro das diversas áreas que envolve uma exploração mineral. Portanto, se faz necessário pesquisar e investir em meios que tornem as atividades mineiras mais ótimas, sustentáveis e econômicas. Este trabalho irá discursar sobre o uso de uma nova ferramenta chamada de *Big Data analytics* nas atividades mineiras e também apresentar o resultado da sua aplicação no processamento mineral de uma mina de ouro localizada na Austrália.

Palavras-chave: produtividade, *Big Data*, *blending*, processamento mineral.

ABSTRACT

The success of mining site is crucially related to mining site productivity which means in how many workers, equipments and materials are used to produce a specific ore quantity. To achieve a satisfactory productivity, it is necessary produce and to market highly attending all specifications market about ore grade. This grade ore is directly related to concentration of ore in final product after mineral processing and many times it does not achieve costumers' requirements. Usually, to fix it the mining companies mixes the low-grade ore with high-grade ore to improve the quality and achieve the correct specifications market, this process is called *blending*. Recently, the mining activities has confronted some challenges related to productivity because increasingly the mining companies has worked with low grade ore due to resource depletion. Then, to maintain a high production with acceptable ore specifications, the mining companies has invested in new technologies to optimizing several activities which belongs to mineral extraction. Therefore, it is essential research and invest in mechanism that become mining activities more optimum, sustentable and economy. This work will discuss about the use of Big Data analytics in mining activities and also present the result of its application in mineral processing plant in a gold mine site located in Australia.

Key words: productivity, Big Data, blending, mineral processing.

LISTA DE FIGURAS

FIGURA 1: MORTES EM ACIDENTES NAS MINAS DOS ESTADO UNIDOS ENTRE 1910 E 2000.	13
FIGURA 2: TRATAMENTO E TRANSFORMAÇÃO DOS DADOS EM FORMATOS ADEQUADOS PARA O PROCESSO DE DATA MINING. 21	
FIGURA 3: ÁRVORE DE DECISÃO ORIUNDA DOS DADOS DA TABELA 2.....	23
FIGURA 4: AGRUPAMENTO DE OBJETOS EM TRÊS GRUPOS USANDO O <i>K-MEANS</i>	24
FIGURA 5: PAINEL ESTRUTURAL DE UM SISTEMA DE BIG DATA ANALYTICS	28
FIGURA 6: SISTEMA DE GERENCIAMENTO DE FROTAS	29
FIGURA 7: CONSUMO DE REAGENTE MEDIDO X CONSUMO DE REAGENTE PREVISTO – DADOS DE LIXIVIAÇÃO	40
FIGURA 8: VALORES ABSOLUTOS DOS COEFICIENTES DE CADA VARIÁVEL DE ENTRADA – REGRESSÃO LINEAR - DADOS DE LIXIVIAÇÃO	41
FIGURA 9: CONSUMO DE REAGENTE MEDIDO X CONSUMO DE REAGENTE PREVISTO – DADOS DA LIXIVIAÇÃO	42
FIGURA 10: IMPORTÂNCIA DAS VARIÁVEIS ATRAVÉS DA DIFERENÇA ENTRE R^2 E R_{x_i} – DADOS DA LIXIVIAÇÃO.....	43
FIGURA 11: PERDA DE OURO NÃO-DISSOLVIDO MEDIDO X PERDA DE OURO NÃO-DISSOLVIDO PREVISTO	44
FIGURA 12: VALORES ABSOLUTOS DOS COEFICIENTES DE CADA VARIÁVEL DE ENTRADA – DADOS DA PERDA DE OURO – REGRESSÃO LINEAR	45
FIGURA 13: CONSUMO DE REAGENTE MEDIDO X CONSUMO DE REAGENTE PREVISTO – PERDA DE OURO – <i>RANDOM FOREST</i> 46	
FIGURA 14: IMPORTÂNCIA DAS VARIÁVEIS ATRAVÉS DA DIFERENÇA ENTRE R^2 E R_{x_i} - PERDA DE OURO – <i>RANDOM FOREST</i>	47

LISTA DE TABELAS

TABELA 1: ÍNDICE DE DESENVOLVIMENTO HUMANO POR REGIÃO	12
TABELA 2: DADOS SOBRE 860 PESSOAS RECÉM FALECIDAS PARA O ESTUDO DOS EFEITOS DA BEBIDA, CIGARRO E EXCESSO DE PESO NA EXPECTATIVA DE VIDA.	22
TABELA 3: DADOS DE LIXIVIAÇÃO DA PLANTA DE BENEFICIAMENTO DE OURO	36
TABELA 4: REPRESENTATIVIDADE DE CADA VARIÁVEL DOS DADOS DE LIXIVIAÇÃO NA PLANTA DE BENEFICIAMENTO DE OURO .	36
TABELA 5: REPRESENTATIVIDADE DAS VARIÁVEIS DOS DADOS DA PERDA DE OURO NA PLANTA DE BENEFICIAMENTO.	39
TABELA 6: VALORES DE R^2 PARA OS MODELOS DE PREVISÃO CONFORME A AUSÊNCIA DE CADA VARIÁVEL – DADOS DA LIXIVIAÇÃO	41
TABELA 7: VALORES DE R_{x_i} PARA OS MODELOS <i>RANDOM FOREST</i> CONFORME A AUSÊNCIA DE CADA VARIÁVEL – DADOS DE LIXIVIAÇÃO	43
TABELA 8: VALORES DE R^2 PARA OS MODELOS DE REGRESSÃO LINEAR CONFORME A AUSÊNCIA DE CADA VARIÁVEL – DADOS DA PERDA DE OURO	44
TABELA 9: VALORES DE R_{x_i} PARA OS MODELOS <i>RANDOM FOREST</i> CONFORME A AUSÊNCIA DE CADA VARIÁVEL – DADOS DA PERDA DE OURO	46

SUMÁRIO

1	INTRODUÇÃO	7
1.1	OBJETIVOS	8
1.1.1	<i>Objetivo Geral</i>	8
1.1.2	<i>Objetivos Específicos</i>	8
2	METODOLOGIA	9
2.1	REVISÃO BIBLIOGRÁFICA	9
2.2	ESTUDO DE CASO	9
3	INDÚSTRIA MINERAL	11
3.1	A IMPORTÂNCIA DA INDÚSTRIA MINERAL	11
3.2	A NECESSIDADE DE NOVAS TECNOLOGIAS	12
3.2.1	<i>Tecnologias implantadas</i>	14
4	BIG DATA	17
4.1	O QUÊ É BIG DATA?	17
4.1.1	<i>Características do Big Data</i>	17
4.1.2	<i>Machine Data</i>	18
4.2	BIG DATA ANALYTICS	18
4.2.1	<i>Machine learning (Aprendizagem de Máquina)</i>	20
4.2.2	<i>Data mining (Mineração dos dados)</i>	21
4.2.3	<i>Classification and regression (Classificação e regressão)</i>	22
4.2.4	<i>Decision Tree (Árvore de decisão)</i>	23
4.2.5	<i>Clustering (Agrupamento)</i>	24
4.2.6	<i>Pattern Recognition (Reconhecimento de padrões)</i>	25
4.2.7	<i>Anomaly Detection (Detecção de anomalias)</i>	25
5	BIG DATA ANALYTICS NA MINERAÇÃO	26
5.1	A IMPORTÂNCIA DO BIG DATA ANALYTICS NAS ATIVIDADES MINEIRAS	26
5.2	O FUNCIONAMENTO DESTA TECNOLOGIA NA MINERAÇÃO	27
5.2.1	<i>Sensores</i>	27
5.2.2	<i>Fontes dos dados</i>	27
5.2.3	<i>Tomada de decisões</i>	27
5.2.4	<i>Estrutura de um sistema de Big Data Analytics</i>	28
5.3	CASOS DE SUCESSO DO USO DE BIG DATA ANALYTICS NA MINERAÇÃO	32
5.3.1	<i>A internet das coisas</i>	32
5.3.2	<i>Veículos inteligentes</i>	33
5.3.3	<i>Mídia Social e contribuição colaborativa (Crowdsourcing)</i>	33

6	MINA DE OURO NA AUSTRÁLIA	35
6.1	METODOLOGIA ANALÍTICA	35
6.1.1	<i>Regressão Linear</i>	19
6.1.2	<i>Random Forest</i>	19
6.2	ESTUDO E ANÁLISE DOS DADOS	35
6.2.1	<i>Dados de lixiviação da planta de beneficiamento de ouro (LeachData)</i>	35
6.2.2	<i>Dados da perda de ouro (Goldloss data)</i>	38
6.3	RESULTADOS	40
6.3.1	<i>Dados de lixiviação</i>	40
6.3.2	<i>Dados da perda de ouro</i>	44
7	CONCLUSÃO	48
	REFERÊNCIAS BIBLIOGRÁFICAS	49
	APÊNDICE A – DADOS DA PERDA DE OURO	52
	APÊNDICE B – MODELO COMPLETO DA REGRESSÃO LINEAR PARA OS DADOS DE LIXIVIAÇÃO	53
	APÊNDICE C – COMANDOS USADOS PARA O MODELO RANDOM FOREST NO MATLAB	54
	APÊNDICE D - MODELO COMPLETO DA REGRESSÃO LINEAR PARA OS DADOS DA PERDA DE OURO....	55

1 INTRODUÇÃO

A mineração é uma das atividades mais remotas da história da humanidade, desde o começo da civilização, o homem tem usado ferramentas provenientes da superfície terrestre, como as pedras e os metais. O ato de minerar está tão intrinsicamente relacionado ao desenvolvimento da espécie humana que os historiadores categorizam os períodos arqueológicos como a Era da Pedra, do Bronze e do Ferro. Cada passo dado nesta linha evolucionária só foi possível devido a invenção de novas ferramentas que habilitavam os nossos ancestrais a escavar mais profundamente e extrair novos metais e minerais. Apesar de ter passado por diversas transformações durante a evolução humana, a atividade mineira ainda carrega velhos costumes e, historicamente, abrange grandiosos volumes de investimentos e receitas. Recentemente, os preços das commodities vem caindo drasticamente, além disso, as minas com altos teores de minério têm se tornadas escassas e a segurança das atividades, que antes detinha pouca importância, vem se tornando o centro das atenções. Assim, as companhias mineradoras necessitam responder a este quadro de desafios e mudar seu modo de exploração atendendo as características do mercado atual.

A resposta para estes novos desafios pode estar em novas tecnologias como a análise e o uso de *Big Data* que otimizam e diminuem os custos operacionais de um empreendimento minerário. A uso de uma enorme quantidade de dados é uma tendência que muitas empresas estão utilizando para melhorar os seus processos e construir soluções inovadoras para obter espaço neste mercado cada vez mais competitivo. Este tipo de tecnologia pode proporcionar uma clara e ampla visão dos processos para os engenheiros (Refermat, 2015). Oracle e Taylor (2014 *apud* Farah, 2016) argumentam que a coleta e armazenamento de dados podem ser analisados de forma a criar relatórios de acordo com as exigências e necessidades da companhia, além disso, o estudo destes dados pode acarretar em descobertas que podem ser lucrativas para a empresa.

Este presente trabalho apresenta o quê é *Big Data*, o porquê do uso desta tecnologia na mineração, como é feita a armazenagem, análise e aplicação destes dados e os métodos utilizados para tal. Além disso, apresenta um estudo de caso feito em uma mina de ouro na Austrália.

1.1 Objetivos

1.1.1 Objetivo Geral

Estudo da análise e utilização de enorme quantidade de armazenamento de dados provenientes dos equipamentos utilizados na extração mineral objetivando a otimização da cadeia produtiva e também a redução de gastos e materiais.

1.1.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

1. Realização de revisão bibliográfica sobre a indústria mineral, *Big Data* e aplicações destas ferramentas em minas existentes.
2. Realização de estudo de caso em uma planta de beneficiamento de ouro na Austrália como exemplo da aplicação de análise de dados na mineração.
3. Construção de dois modelos preditivos: regressão linear e *random forest* para explicar o comportamento da variável de saída em relação as outras variáveis de entrada dos modelos.
4. Avaliação do comportamento de cada variável de entrada nos modelos preditivos utilizados.

2 METODOLOGIA

A metodologia utilizada para realização deste trabalho será dividi-lo em duas partes principais: revisão bibliográfica e estudo de caso.

2.1 Revisão Bibliográfica

A revisão bibliográfica consiste em pesquisar e coletar dados sobre os seguintes assuntos:

- Importância da indústria mineral: o papel da mineração na sociedade atual e os aspectos que levaram o desenvolvimento de mecanismos que ajudaram o progresso da mineração ao longo da história.
- Tecnologias usadas na mineração: o porquê do desenvolvimento de novas tecnologias e seus respectivos benefícios.
- *Big Data*: o que é *Big Data*, as características do *Big Data*, a estrutura desta tecnologia e os dados de máquina.
- *Big Data Analytics*: aprendizado de máquina, mineração de dados, classificação e regressão, árvore de decisão, *clustering*, reconhecimento de padrões e detecção de anomalias.
- Exemplos do uso de *Big Data* na mineração: casos de sucesso em minas em operação que utilizam a análise de dados para otimizar seus processos.

2.2 Estudo de caso

Estudar o caso consiste em apresentar e analisar os principais elementos associados ao uso de mecanismos relacionados ao *Big Data analytics* na melhoria da planta de beneficiamento de uma mina de ouro na Austrália.

- Dados de lixiviação: estudo e análise dos dados de lixiviação da planta e resultados destes dados após o devido tratamento.
- Dados da perda de ouro: estudo e análise dos dados referentes à perda de ouro no tratamento deste minério e os resultados após o tratamento destes dados.
- Resultados finais: resumo dos resultados obtidos quando os modelos de correção são aplicados nos dois casos acima. Resultados sobre os efeitos de diferentes manuseios

das variáveis de saída e quais variáveis influenciam mais na planta de beneficiamento de ouro.

3 INDÚSTRIA MINERAL

3.1 A importância da Indústria Mineral

Desde o início da civilização, a humanidade tem utilizado métodos de mineração para extrair minerais da superfície terrestre. Sociedades antigas do Egito, Roma e Grécia já eram engajados e detinham o conhecimento e as ferramentas para minerar. Fundamentalmente relacionado à evolução da sociedade, a mineração nos tempos atuais, representa, além de uma fonte de minerais minério, uma manutenção e melhoria dos padrões de vida de um país (NRC, 2002).

Os materiais oriundos da mineração são necessários para a construção de estradas, hospitais, automóveis e casas. Além disso, são essenciais para a produção de computadores, satélites e eletricidade e proporcionam vários bens de consumo e serviços para os consumidores (NRC, 2002). Uma mina em atividade pode ser considerada economicamente importante para a produção da região, pois provém empregos e impostos que são investidos em hospitais, escolas e políticas públicas para a melhoria da qualidade de vida da comunidade local. As atividades minerais também concedem cursos profissionalizantes e treinamentos adequados aos trabalhadores e fomentam o comércio regional devido à necessidade de pequenos serviços (NRC, 2002).

De acordo com o Conselho Nacional de Pesquisa dos Estados Unidos (NRC), as mineradoras possuem papel crucial ao produzir rendimentos que são negociados internacionalmente contribuindo para o crescimento do produto interno bruto de um país. No Brasil, de acordo com o Instituto Brasileiro de Mineração (IBRAM), em 2014, a mineração correspondeu a 5% do PIB do país com um valor de US\$40 bilhões. Outro aspecto importante é a contribuição da mineração no crescimento do IDH de uma região, a Tabela 1 apresenta este crescimento em algumas regiões do Brasil onde há produção mineral. A mineração também promove uma gama de atividades associadas, como por exemplo, a fabricação de equipamentos de mineração, serviços de engenharia e preservação ambiental e o desenvolvimento de universidades nos campos de geologia, mineração e metalurgia. Conclui-se, portanto, que as oportunidades econômicas e riquezas geradas por esta atividade são substanciais para os países produtores (NRC, 2002).

Tabela 1: Índice de Desenvolvimento Humano por região

Município	IDH município	IDH estado
Nova Lima (MG)	0.813	0.731
Parauapebas (PA)	0.715	0.646
Barro Alto (GO)	0.742	0.735
Ariquemes (RO)	0.702	0.690
Criciúma (SC)	0.788	0.774

Fonte: Instituto Brasileiro de Mineração (IBRAM)

3.2 A necessidade de novas tecnologias

Ao longo do século passado, a indústria mineral cresceu vagarosamente até 1980s quando os preços das commodities estagnou, permanecendo assim até os anos 90's. Após este período, houve um crescimento nos preços e também em toda a indústria extrativista devido ao crescimento da construção civil fomentado pela demanda da China e outros países em desenvolvimento. Este rápido crescimento foi chamado de “super ciclo” e por se tratar de um ciclo tornaria a decrescer. Desde 2011, os preços das commodities voltaram a cair novamente, forçando as mineradoras a lidar com esta nova tendência de maneira rentável e inovadora, mesmo que alguns afirmassem ser o fim da idade do ferro (Lee, 2014).

De acordo com Lee (2014), por se tratar de uma atividade antiga que carrega métodos tradicionais de operação, qualquer mudança nas atividades mineiras tem sido lenta e gradual. As companhias mineradoras são muitas vezes relutantes em abraçar novas tecnologias, ou seja, é muito difícil que elas deixem de trabalhar com processos já testados e aprovados quando o foco é apenas a produção e o lucro. Usualmente, a adoção de uma nova tecnologia, assim como qualquer novo processo, apenas é considerada quando provém melhorias em um processo antigo ou atende à um requisito exigido (Lee, 2014).

Outro importante aspecto a se considerar em uma atividade minerária é a segurança e a saúde dos trabalhadores. No começo do século vinte, nos Estados Unidos, aproximadamente 3.000 trabalhadores de minas de carvão e 1.000 trabalhadores de minas de minerais não-metals eram mortos anualmente em acidentes (Ramani and Mutmansky, 2000 *apud* Lee, 2014). Ainda se

tratando deste mesmo país, depois da criação do Escritório de Minas dos Estados Unidos (USBM) em 1910, houve um enorme esforço em pesquisa e identificação dos perigos relacionados à mineração. Desde aquele tempo, as condições de trabalho têm melhorado em consequência de avanços no desenvolvimento e design da mina, da rigorosa legislação sobre saúde e segurança e também da introdução de sistemas mais produtivos. Houve também trabalhos em conjunto entre a indústria mineral e o USBM que permitiram um progresso na ciência e tecnologia relacionadas à mineração reduzindo drasticamente as mortes e acidentes (Ramani and Mutmansky, 2000 *apud* Lee, 2014). Na Figura 2, é possível constatar exatamente este decréscimo de fatalidades após o processo de implementação de novas tecnologias.

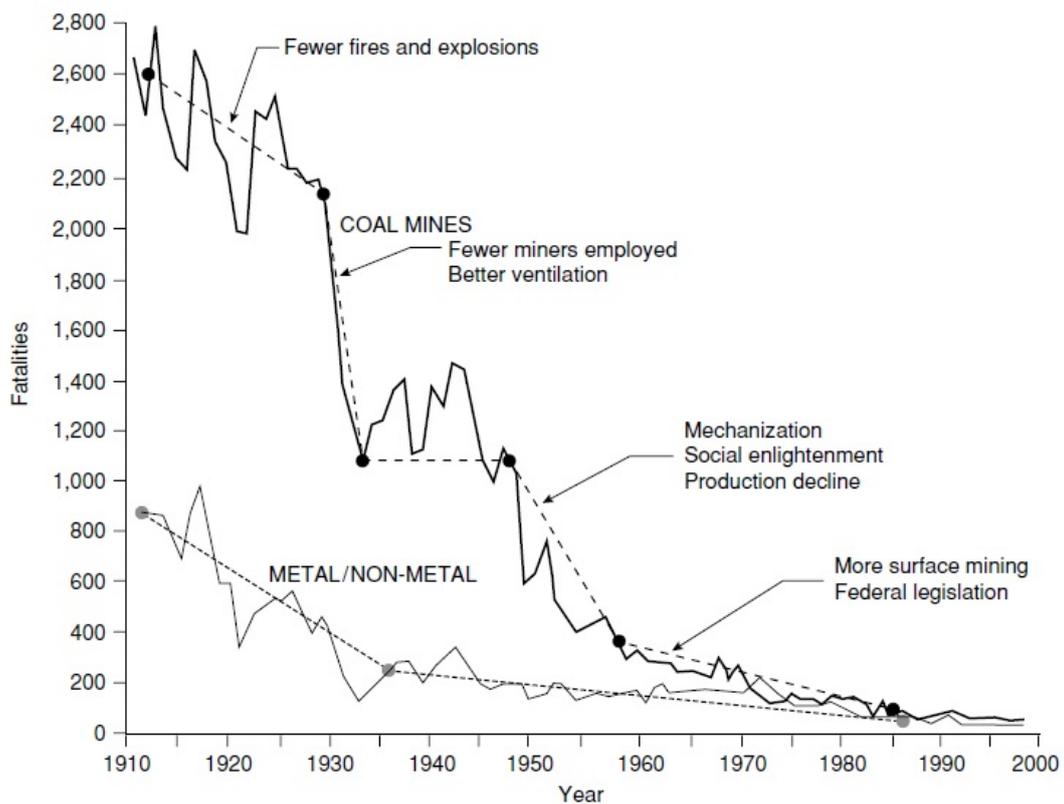


Figura 1: Mortes em acidentes nas minas dos Estado Unidos entre 1910 e 2000.

Fontes: Katen, 1992; Ramani and Mutmansky, 2000; U.S. Department of Labor, 1999.

Diante do cenário econômico adverso e considerando a imensa preocupação dada à segurança e saúde dos trabalhadores, as empresas mineradoras necessitam mudar a forma como trabalham e adotar novas formas de extração. Embora algumas delas permaneçam relutantes,

elas não devem fugir dos benefícios e possibilidades provenientes do uso de novas metodologias, como por exemplo a aplicação da análise de *Big Data*, elas entendem que a inovação deve ser implementada para que os equipamentos usados na mina trabalhem mais eficientemente e proporcionem mais segurança aos trabalhadores (LEE, 2014).

3.2.1 Tecnologias implantadas

O emprego de tecnologias e metodologias nas atividades que englobam uma mina, além de proporcionar eficiência nas operações de exploração, lavra e processamento, também estão preocupadas em garantir mais segurança e saúde para os trabalhadores. Esta seção irá discutir algumas mudanças e a utilização destas tecnologias na forma de minerar. As áreas abordadas serão: tamanho dos equipamentos, automação, fontes alternativas de energia, comunicação e treinamento tecnológico.

3.2.1.1 Tamanho dos equipamentos

Uma tendência da indústria mineral é o constante aumento do tamanho das máquinas utilizadas na mina. Conforme há o crescimento do nível de automação, enormes equipamentos podem estar aptos a operar em um modo totalmente automatizado (Shuey, 1999 *apud* NRC, 2002). O Conselho Nacional de Pesquisa dos Estados Unidos argumenta que grandes equipamentos certamente evita a exposição individual dos trabalhadores ao ambiente mineiro. Entretanto, no caso de equipamentos de carregamento, a visibilidade é comprometida aumentando as chances de ocorrer um acidente. Sendo assim, a tecnologia implementada para solucionar este caso é o uso de sensores que avisam a presença de obstruções, outros equipamentos, bermas e trabalhadores. Outra tecnologia usada para a segurança e eficiência é o uso de equipamentos automáticos guiados por sensores internos e GPS (NRC, 2002).

3.2.1.2 Automação

NIOSH (1999 *apud* NRC, 2002) argumenta que muitos métodos de lavra, como o *Longwall* em minas subterrâneas de carvão, já são altamente automatizados. Semi autônomos e autônomos sistemas resultam em ótimos resultados de produção e produtividade, bem como em melhores condições de saúde e segurança. Apesar disso, máquinas automatizadas mesmo que controlados por equipamentos remotos podem criar perigos, pois o número de sistemas

automatizados e a disponibilidade de dados sobre segurança e saúde são ainda pequenos. Entretanto, é fato que a automação tem sido responsável por um índice pequeno de acidentes na mineração envolvendo mortes e incapacidades (NIOSH, 1999, *apud* NRC, 2002).

3.2.1.3 Fontes alternativas de energia

As fontes alternativas de energia têm sido estudadas e implementadas por aumentar a eficiência da atividade minerária e por reduzir os riscos e perigos aos trabalhadores. De acordo com NRC (2002), as fontes como o hidrogênio tem sido desenvolvida, mas os equipamentos movidos a diesel ainda são os mais usados, apesar da segurança nas operações com este tipo de máquina ser afetada por gases, poeira, visibilidade, vibrações e ruídos. Nas minas subterrâneas, os equipamentos a diesel estão sendo substituídos pelas máquinas elétricas e isto se deve à necessidade e o custo de se ter uma grande estrutura de ventilação que mantenha o ambiente seguro e saudável, ou seja, livre dos poluentes gerados pela combustão do diesel. Além de prover uma atmosfera limpa, a energia consumida pelos caminhões é reduzida em 70% devido ao uso de motores elétricos (Jensen, 2013).

3.2.1.4 Comunicação

Pesquisas e estudos que desenvolvem mecanismos de monitoramento computadorizado sobre as condições da mina com um sistema de informações seguro e rápido e que transmita estas informações em tempo real para os técnicos podem ter excelentes ganhos para a indústria mineral (NRC, 2002). Devido a alta automatização dos equipamentos e sistemas utilizados na mina, os técnicos costumam não trabalhar próximos ou não ter constante contato visual uns com os outros. De fato, muitos destes trabalhadores exercem suas funções em áreas remotas da mina. Conseqüentemente, a necessidade de comunicar uns com os outros baseando-se em dados e análises em tempo real pode ser comprometida. Sendo assim, componentes que monitoram a atmosfera e sistemas que monitoram os equipamentos tem sido desenvolvidos e implementados. Estes avanços em comunicação devem ser integrados especificamente para que as informações sobre segurança e instruções possam ser compartilhadas em tempo real pelos técnicos da mina (NRC, 2002).

3.2.1.5 Treinamento tecnológico

Qualquer mudança ou instalação de uma nova tecnologia ou metodologia nos processos e atividades da mina requer um treinamento adequado para os trabalhadores. Situações com um novo ambiente, novo equipamento, novas regras ou uma nova equipe de técnicos devem ser submetidas a este tipo de capacitação (NRC, 2002). O Conselho Nacional de Pesquisa dos Estados Unidos argumenta que os treinamentos obrigatórios de saúde e segurança podem ser aperfeiçoados por ferramentas de treinamentos orientadas por computador, as capacitações tecnológicas são capazes de usar a habilidade dos computadores de superar as limitações dos tradicionais treinamentos em salas de aula e no trabalho. As capacitações baseadas na informática também promovem melhores projetos, alta qualidade de lições e auto aprendizado, utilizam módulos de treinamentos com realidade virtual resultando na melhoria da capacidade de os trabalhadores reagirem a uma determinada situação de perigo (NRC, 2002). Um exemplo deste tipo de capacitação é o treinamento dos motoristas de caminhões fora de estrada em simuladores que proporcionam uma realidade virtual da mina.

4 BIG DATA

4.1 O quê é Big Data?

O termo “Big Data” quando traduzido para a língua portuguesa significa “Grandes Dados” e, resumidamente, é exatamente este o conceito deste termo. *Big Data* é a expressão usada para descrever o imenso volume de dados que são gerados diariamente. Em 2012, por dia, foram gerados em torno de 2,5 quintilhões de bytes de dados e 90% destes dados foram produzidos nos dois últimos anos anteriores (IBM, 2012). De acordo com a IBM (2013), este número de dados pode chegar a 2,3 trilhões de gigabytes diários até 2020. Estas informações são oriundas de todo e qualquer lugar, alguns dos exemplos são: sensores usados para reunir informações climáticas, postagens em redes sociais, vídeos e fotos digitais, transações econômicas e sinais de GPS para celular (IBM, 2012).

É quase inimaginável a quantidade de informações que está sendo gerada e armazenada todos os dias em escala mundial. Apesar deste gigantesco volume de material, o *Big Data* não se trata somente da quantidade de informações que estes dados representam, mas também na oportunidade em encontrar intuições ou tendências em um novo e emergente tipo de conteúdo, ou seja, torna possível responder questões anteriormente consideradas além do alcance (IBM, 2012). De acordo com Hekima (2017), todas estas informações representam para as empresas a chance de tomar decisões mais precisas e confiáveis objetivando a otimização e o sucesso do empreendimento.

4.1.1 Características do Big Data

De acordo com a IBM (2012), o processo de construção do *Big Data* abrange três dimensões: volume, velocidade e variedade.

- Volume: refere-se à quantidade de informações produzidas. As empresas são diariamente inundadas com dados cada vez maiores de todos os tipos acarretando no acúmulo de terabytes ou até mesmo pentabytes de informações.
- Velocidade: os dados são gerados rapidamente e devem ser tratados em tempo hábil. Para processos sensíveis ao tempo, como a captura de fraude, a enorme quantidade de dados deve ser usada com agilidade e fluidez.

- Variedade: o processo de geração do *Big Data* deriva de qualquer tipo de dado como textos, informações de sensores, áudio, vídeo e outros.

Ohlhorst (2012) introduz mais uma dimensão pertencente as características do *Big Data*: a veracidade.

- Veracidade: esta dimensão refere-se a anomalias, ruídos e erros estatísticos nos dados produzidos. Quando há milhões de informações armazenadas, torna-se inevitável que algumas destas informações não sejam irrelevantes ou contenham erros resultando em uma possível falsa interpretação sobre os dados. Alguns exemplos são a imprecisão de sensores de medição e a falta de credenciais nas mídias sociais (Ohlhorst, 2012).

4.1.2 *Machine Data*

As fontes de geração de *Big Data* são comumente identificadas como websites, mídias sociais, documentos e outros, mas não estão limitadas apenas a geração por atividades humanas. Uma significativa quantidade de informações também é criada por sistemas e máquinas provindo informações que também podem ser usadas para a otimização do processo, redução de custos e economia de materiais. Este tipo de dado é considerado mais organizado e não segue a real desordem do mundo e sim uma estrutura pré-definida. Os dados provenientes de máquinas são gerados em todas as indústrias, desde equipamentos de saúde até dispositivos portáteis e podem ser usados para encontrar padrões e ou prever tendências do mercado. Além disso, possuem vários casos de uso comercial, como análise de desempenho, previsões de mercado e detecção de fraudes (Surange & Bansal, 2013).

4.2 *Big Data Analytics*

O sistema de *Big Data Analytics* visa encontrar soluções que permitam: redução de custos; redução de tempo; desenvolvimento de novos produtos e decisões mais inteligentes. Para isto, este processo baseia-se em um minucioso trabalho de análise de grandes volumes de dados que são coletados, armazenados e interpretados por softwares de altíssimo desempenho. Consiste do cruzamento de uma imensidade de dados do ambiente interno e externo resultando em um sistema de prognósticos para tomadores de decisões (Hekima, 2017). Russom (2011) argumenta que a análise de grandes dados ajuda os interessados em descobrir

o quê tem mudado e como eles devem reagir a esta mudança. À medida que o mercado mundial avança para a recuperação, há cada vez mais oportunidades de negócios que devem ser aproveitadas e a análise dos dados é a maneira mais eficiente de descobrir novos segmentos de clientes, identificar os melhores fornecedores e associar produtos com a afinidade do mercado (Russom, 2011). Portanto, o uso do *Big Data Analytics* está se tornando uma base fundamental para a concorrência e crescimento das empresas e provavelmente fomentará nova ondas de produtividade, crescimento e superávit de consumo (Ohlhorst, 2012).

A seguir, se faz necessário descrever os aspectos que constituem o sistema de *Big Data Analytics* justificado pela tamanha importância e papel que esta análise possui como fator determinante no sucesso das inferências e previsões.

4.2.1 Regressão Linear

O modelo de regressão linear baseia-se em uma variável dependente ou variável de interesse em função de outras variáveis que são independentes. A relação entre estas variáveis é o centro da regressão linear e através desta relação pode ser previsto o comportamento da variável dependente. Quando um conjunto de dados possui mais de uma variável independente, o modelo é chamado de regressão linear múltipla. No caso do processamento mineral, a regressão linear múltipla pode prever a concentração de minério e a recuperação da planta de beneficiamento (Wills, 2013).

4.2.2 Random Forest

O modelo de *random forest* é um modelo não linear de aprendizagem supervisionada que utiliza da árvore de decisões para parametrizar os dados. O modelo pode ser considerado incrivelmente preciso entre os algoritmos atuais e pode ser executado com competência em grandes bases de dados. Como modelo preditivo, o *random forest* pode contribuir para otimizar o processo tanto na mineração como na metalurgia do minério. Em alguns casos, no processamento mineral, quando a relação das variáveis é não-linear ou complexa, pode ser interessante usar modelos não-lineares, como o *random forest*, para alcançar um programa seguro (Auret & Aldrich, 2012).

4.2.3 *Machine learning* (Aprendizagem de Máquina)

A Aprendizagem de Máquina tem como objetivo elaborar algoritmos e métodos que possibilitem a um computador “aprender”, isto é, otimizar o seu desempenho em uma determinada tarefa ou na solução de um problema. Todo e qualquer problema computacional é devidamente solucionado através de um algoritmo. Um algoritmo pode ser conceituado como uma sequência de instruções que devem ser efetuadas para transformar os dados de entrada em dados de saída. Para elucidar estes conceitos, tomemos como exemplo um algoritmo para fazer uma triagem qualquer. Neste caso, os dados de entrada é um conjunto de números e os dados de saída é uma lista ordenada destes números. Para a mesma tarefa, existem vários algoritmos e o ideal é o descobrimento do mais eficiente deles exigindo o menor número de instruções ou memória (Alpaydín, 2010).

Machine learning ou aprendizagem das máquinas respalda-se na programação de computadores para melhorar a performance de um processo criterioso usando exemplos de dados ou experiências passadas. De posse de um modelo definido para alguns parâmetros, a aprendizagem baseia-se na execução de um programa computacional para otimizar os parâmetros do modelo usando os dados de teste ou experiências anteriores. O modelo pode ser preditivo, fazendo previsões sobre o futuro, descritivo, obtendo conhecimento acerca dos dados, ou ambos (Alpaydín, 2010).

De acordo com Alpaydín (2010), a aprendizagem das máquinas trabalha com métodos estatísticos na construção de modelos matemáticos, pois o objetivo principal deste processo consiste em fazer inferências acerca de uma amostra. Para tal, o papel da ciência dos computadores pode ser considerado de duas formas:

1. Treinamento: é necessário que haja algoritmos eficientes para resolver os problemas de otimização, armazenamento e processo desta grande quantidade de dados produzidos.
2. Representação: após o aprendizado de um modelo, se faz necessário a representação deste e uma solução algorítmica para alcançar eficientes e precisas inferências.

Sendo assim, as técnicas que envolvem a aprendizagem de máquina não revelam os motivos de um determinado evento ou acontecimento, entretanto, podem construir modelos a partir dos grandes dados sobre as correlações que podem ser usadas para calcular os resultados (Alpaydín, 2010).

4.2.4 *Data mining* (Mineração dos dados)

Data mining pode ser definido como o processo de descoberta de padrões e preferências contidos nos dados. Este processo é um componente do *Machine learning* e seu objetivo principal é procurar relações insuspeitas e apresentar as informações de maneira inovadora, de tal forma que as informações sejam compreensíveis e úteis para o proprietário dos dados (Hand, Mannila, & Smyth, 2001). Como os dados são originados de diferentes fontes, eles podem ser estruturados ou não. Assim, para a implementação das técnicas de *data mining* é preciso primeiro que estes dados sejam transformados em um lógico e organizado formato, antes que os mesmos sejam usados como dados de entrada para a análise (Reffat, Gero, & Peng, 2004). A Figura 2 resume este método.

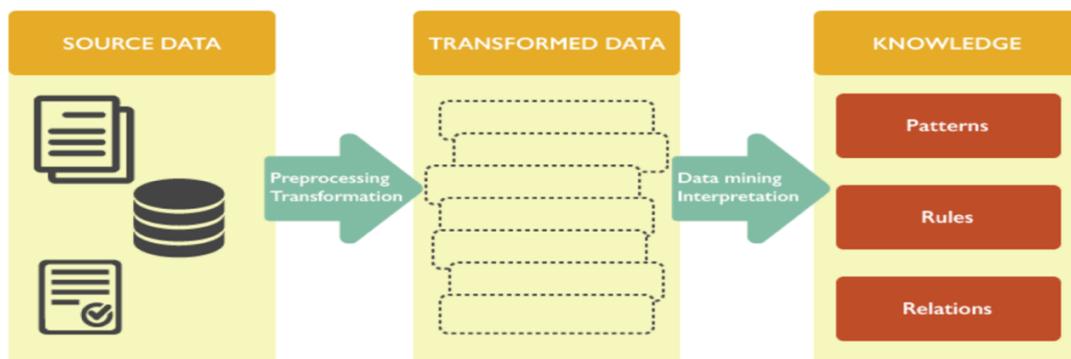


Figura 2: Tratamento e transformação dos dados em formatos adequados para o processo de data mining.

Fonte: Adaptado de Reffat, R.M., Gero, J.S., & Peng, W. (2004)

Existem diversas técnicas de análise na mineração de dados baseadas no propósito da pesquisa. Estas técnicas são classificadas em aprendizagem supervisionada e aprendizagem

não supervisionada. A primeira assume dados rotulados, ou seja, cada objeto é rotulado por uma variável de resposta. O objetivo desta técnica é usar os dados de entrada para prever os valores dos dados de saída. A aprendizagem não supervisionada é efetuada com base em observação e descoberta, isto é, os dados não são rotulados. A tarefa deste método consiste em descrever como os dados são organizados ou agrupados. (van der Aalst, 2011).

4.2.5 *Classification and regression (Classificação e regressão)*

Harrison (1998) argumenta que o objetivo da classificação constitui em desenvolver um modelo que possa ser aplicado a dados não classificados visando categorizá-los em classes, ou seja, um objeto é analisado e classificado segundo uma classe definida. As técnicas de classificação assumem uma variável de resposta categórica objetivando classificar uma concreta ocorrência de algum objeto ou característica baseando-se nas variáveis de predição (van de Aalst, 2011).

Para exemplificar esta técnica, observe a Tabela 2 sobre um estudo sobre a relação entre a expectativa de vida e os hábitos das pessoas.

Tabela 2: Dados sobre 860 pessoas recém falecidas para o estudo dos efeitos da bebida, cigarro e excesso de peso na expectativa de vida.

Drinker	Smoker	Weight	Age
Yes	Yes	120	44
No	No	70	96
Yes	No	72	88
Yes	Yes	55	52
No	Yes	94	56
No	No	62	93
...

Fonte: Adaptado de van de Aalst (2011)

O processo consiste em classificar as pessoas na classe de fumantes e na classe de não fumantes. Portanto, o método seleciona uma variável de resposta categórica chamada de *smoker*. Através da classificação, o objetivo é descobrir qual é a diferença chave entre fumantes e não fumantes. Por exemplo, é possível constatar que a maioria dos fumantes que bebem morrem jovens (van de Aalst, 2011).

A regressão tem como objetivo definir um valor para alguma variável contínua desconhecida, as variáveis contínuas geralmente provêm de medições de uma característica do objeto, exemplos: receita, altura ou saldo do cartão de crédito (Harrison, 1988). A regressão assume uma variável de resposta numérica e visa a descoberta de uma função que corresponda aos dados com o menor erro possível. Por exemplo, para os dados da Tabela 2, é possível selecionar a idade como variável de resposta e hipoteticamente encontrar a função:

$$age = (124 - 0,8 \times weight)$$

Aassim, espera-se que uma pessoa com 50kg viva até os 84 anos enquanto que uma pessoa com 100kg viva até os 44 anos (van de Aalst, 2011). Portanto, o papel da regressão é tentar encontrar a função que possa prever as variáveis de resposta com a maior precisão existente com base nos valores de predição.

4.2.6 *Decision Tree* (Árvore de decisão)

A árvore de decisão é usada para construir modelos que fazem previsões sobre específicos valores. Esta técnica baseia-se em diferentes variáveis de predição e visa a classificação de objetos. Nesta técnica existe apenas uma variável de resposta categórica que rotula os dados e o resultado é organizado sob a forma de uma árvore (van de Aalst, 2011). No exemplo apresentado pela Figura 3, encontra-se uma árvore de decisão derivada da Tabela 2.

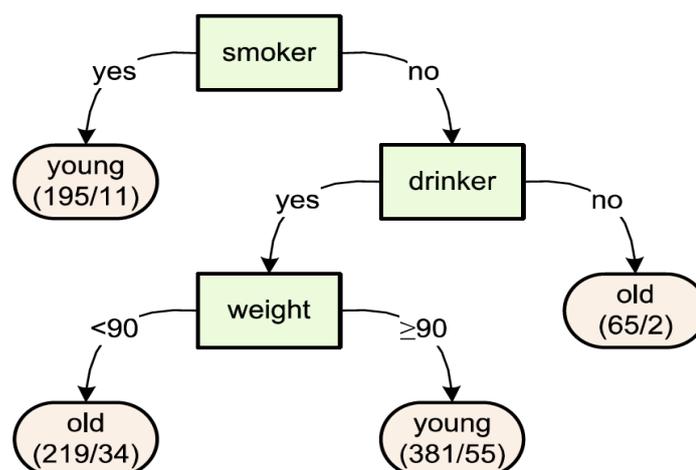


Figura 3: Árvore de decisão oriunda dos dados da Tabela 2
 Fonte: Adaptado de van de Aalst (2011)

A representação se faz de cima para baixo e o primeiro nó ou nó raiz é o grupo de 860 pessoas recém falecidas. Baseado no atributo *smoker* (*fumantes*) estes objetos (pessoas) são divididos em dois grupos: fumantes (195 pessoas) e não fumantes (665 pessoas). O processo de divisão do grupo de fumantes cessa e estes são categorizados como *young* (*jovens*), isto é, fumantes que provavelmente morreram antes dos 70 anos. Os não fumantes são divididos em bebedores e não bebedores. O último grupo é previsto que vivam mais e são classificados como *old* (*velhos*). E desta maneira, o raciocínio segue até a última divisão relacionada ao peso das pessoas. Todos os nós da árvore possuem dois números, o primeiro indica o número de objetos classificadas como tal e o segundo indica o número de casos correspondentes que foram incorretamente classificados. Por exemplo, dos 195 fumantes classificados como *jovens*, 11 pessoas foram incorretamente classificadas, ou seja, não morreram antes dos 70 anos enquanto fumava (van de Aalst, 2011).

4.2.7 Clustering (Agrupamento)

Clustering é um método de *Data Mining* para gerar agrupamentos automáticos de dados segundo seu grau de semelhança. De acordo com van de Aalst (2011), a técnica de *Clustering* baseia-se em agrupamentos de objetos em grupos (*clusters*). Os objetos de um mesmo grupo devem ser similares entre eles e diferentes em relação aos objetos de outro grupo. Além disso, os objetos podem ser analisados por um ou mais atributos e o algoritmo geralmente usado é chamado de *k-means*. Este algoritmo é capaz de distribuir os objetos em um número pré-definido de grupos baseado em suas distâncias Euclidianas. A Figura 4 ilustra a ideia básica de agrupamento.

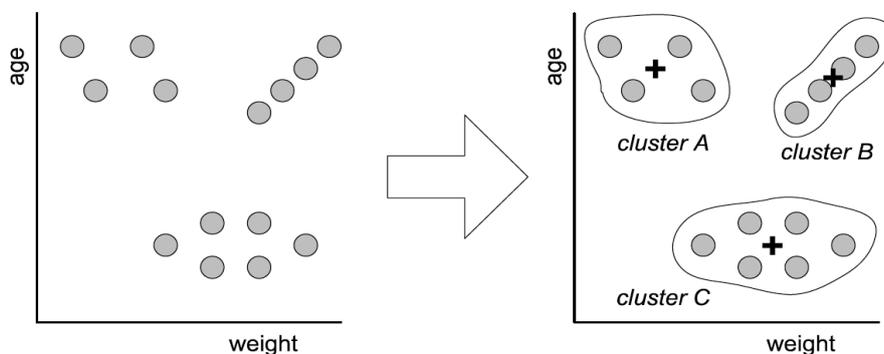


Figura 4: Agrupamento de objetos em três grupos usando o *k-means*
 Fonte: Adaptado de van de Aalst (2011)

Neste exemplo suponha-se ter um conjunto de dados com apenas duas variáveis: idade e peso. Os pontos no gráfico correspondem aos objetos (pessoas) com idade e peso particulares. Através do método de agrupamento *k-means*, os três grupos (clusters) formados no gráfico direito podem ser descobertos. Os objetos de um grupo estão próximos uns dos outros e distantes dos objetos pertencentes aos outros grupos. Cada grupo tem um centro denotado pelo sinal de mais (+) que pode ser calculado pela média das coordenadas das distâncias dos grupos (van de Aalst, 2011).

4.2.8 *Pattern Recognition* (Reconhecimento de padrões)

Bishop (2006) argumenta que a técnica de reconhecimento de padrões fundamenta-se na descoberta automática de regularidades dos dados através do uso de algoritmos computacionais. Por meio destas regularidades ela toma ações como classificar as informações em diferentes categorias. Além disso, este método utiliza diferentes algoritmos para encontrar a característica chave nos dados de entrada. O reconhecimento de padrões pode ser aplicado em conjuntos de dados supervisionados como também nos dados não supervisionados. Como citado na Seção 5.2.2, na aprendizagem supervisionada, o modelo é desenvolvido com um conjunto de dados de teste com objetos adequadamente rotulados e os respectivos dados de saída. Na aprendizagem não supervisionada, os dados de teste não são rotulados, se faz necessário então o uso de diferentes técnicas para identificar padrões nos dados que podem prever o correto valor dos dados de saída (Bishop, 2006).

O reconhecimento de padrões é usado em várias áreas, das quais as mais relevantes são o reconhecimento de fala, reconhecimento óptico, reconhecimento de face e da análise de um relevo em geologia (Bishop, 2006).

4.2.9 *Anomaly Detection* (Detecção de anomalias)

A detecção de anomalias trabalha com o sistema de encontrar padrões nos dados que não apresenta conformidade com os comportamentos esperados. Estes dados em não conformidade são usualmente relacionados às anomalias, valores atípicos, observações discordantes, exceções, aberrações, surpresas e peculiaridades. O uso da técnica de detecção de anomalias possui um amplo campo de aplicação, alguns deles são: detecção de fraudes em cartões de crédito, seguros e assistência médica, detecção de invasões na segurança

cibernética, detecção de falhas em sistemas críticos de segurança e vigilância militar para atividades inimigas (Chandola, Banerjee & Kumar, 2009).

Chandola *et al.* (2009) afirma que a relevância da detecção de anomalias é justificada pelo fato de que as anomalias presentes nos dados representam informações cruciais, e muitas vezes críticas, podendo ser aplicadas em uma variedade de domínios. Por exemplo, um padrão de tráfego anômalo em uma rede de computadores pode significar que um computador hackeado está enviando informações confidenciais para um destino não autorizado (Kumar, 2004, *apud* Chandola *et al.*, 2009). Uma imagem de ressonância magnética que apresente anomalias pode representar a presença de tumores malignos (Spence *et al.*, 2001, *apud* Chandola *et al.*, 2009). Por fim, anomalias nos dados da transação do cartão de crédito podem indicar roubo do cartão ou da identidade (Aleskerov *et al.*, 1997, *apud* Chandola *et al.*, 2009).

5 BIG DATA ANALYTICS NA MINERAÇÃO

5.1 A importância do Big Data Analytics nas atividades mineiras

Para Chakraborti (2016), a aplicabilidade da análise de dados é atualmente uma das grandes questões estudadas pela indústria mineral. De acordo com uma pesquisa feita pelo *Mining Journal*, na primeira metade de 2014, incluindo 10 das 20 maiores mineradoras do mundo, a análise de grandes dados está no topo da lista sobre o quê as empresas vêm como tecnologia capaz de mudar o jogo e transformar a indústria mineral nos próximos cinco a dez anos (Mining Journal, 2015). A análise poderá estimular uma onda de eficiência e ganhos na extração, análise, transporte e processamento do minério, proporcionando mais informações para decisões rápidas e precisas em todos os níveis (Chakraborti, 2016).

Em um mercado cada vez mais competitivo, toda ação para melhorar os processos da cadeia produtiva, usando sistemas inteligentes, é imprescindível. Diante disso, esta é a ação esperada para o *Big Data Analytics* nas atividades mineiras, desempenhando um papel fundamental na condução de uma melhoria na utilização de ativos¹ e no aumento da produtividade (Chakraborti, 2016).

5.2 O funcionamento desta tecnologia na mineração

5.2.1 Sensores

Na base do sistema da análise de dados estão os sensores instalados em todos os dispositivos e equipamentos que fazem parte das atividades de uma mina. Estes sensores, em tempo real, são capazes de gerar dados geocientíficos, de ativos e de dados operacionais. Isto se deve à melhoria da velocidade de troca de dados por Wi-Fi e 3G/4G que proporcionam a coleta de informações em tempo real do ponto de extração do minério até o seu destino final nas plantas de beneficiamento. Estes dados podem ser analisados usando um computador com múltiplos processadores, onde cada processador usa seu próprio sistema operacional e memória. Logo depois, distribui-se rapidamente as informações coletadas para os interessados (Chakraborti,

¹ equipamentos de tecnologia da informação, contratos, equipamentos dos processos de produção, ferramentas e materiais.

2016). Todo este procedimento é possível graças as modernas plataformas de *Big Data* que podem assimilar, em tempo real, imensas quantidades de entradas heterógenas oriundas de múltiplas fontes. Portanto, simultaneamente, as plataformas constroem análises preditivas e prescritivas para impulsionar a excelência operacional (Chakraborti, 2016).

5.2.2 Fontes dos dados

As fontes de dados em uma mina podem ser classificadas como medição direta ou indireta (auxiliar). As fontes de medição direta são aquelas originárias de instrumentos como os convencionais levantamentos topográficos e informações de GPS. As fontes de medição indireta são sistemas que coletam dados como subproduto de processos operacionais, alguns deles são: o sistema de gestão de frotas, o sistema de supervisão e aquisição de dados (SCADA), dados da perfuratriz e dados de modelos geográficos (Chakraborti, 2016).

5.2.3 Tomada de decisões

Chakraborti (2016) cita que para ocorrer uma eficiente extração do corpo mineralizado tornando a recuperação do minério satisfatória, se faz necessário o uso de softwares para a construção de um modelo do corpo do minério. Este modelo fornece dados de padrões

geológicos que determinam os furos de perfuração. Assim, a chave para que as melhores decisões sejam tomadas é a disponibilidade de dados precisos provenientes de múltiplos sistemas combinados com análises simultâneas. Estas decisões podem ser aplicadas à exploração, produção e operações mineiras. Além disso, elas são importantes para detectar e sanar as causas de estrangulamentos operacionais como os atrasos na manutenção dos caminhões, tempo de fila longo de caminhões e LHDs e atrasos em amostras de laboratórios submetidas ao controle de qualidade (Chakraborti, 2016).

5.2.4 Estrutura de um sistema de Big Data Analytics

A Figura 5 é um exemplo de um sistema de análise de grandes dados onde cada camada ou setor exerce uma função. Ela representa as diferentes etapas pelas quais os dados devem passar desde dados estatísticos crus até a inferência de previsões, a primeira etapa deste processo é indicada na base do painel (Chakraborti, 2016).

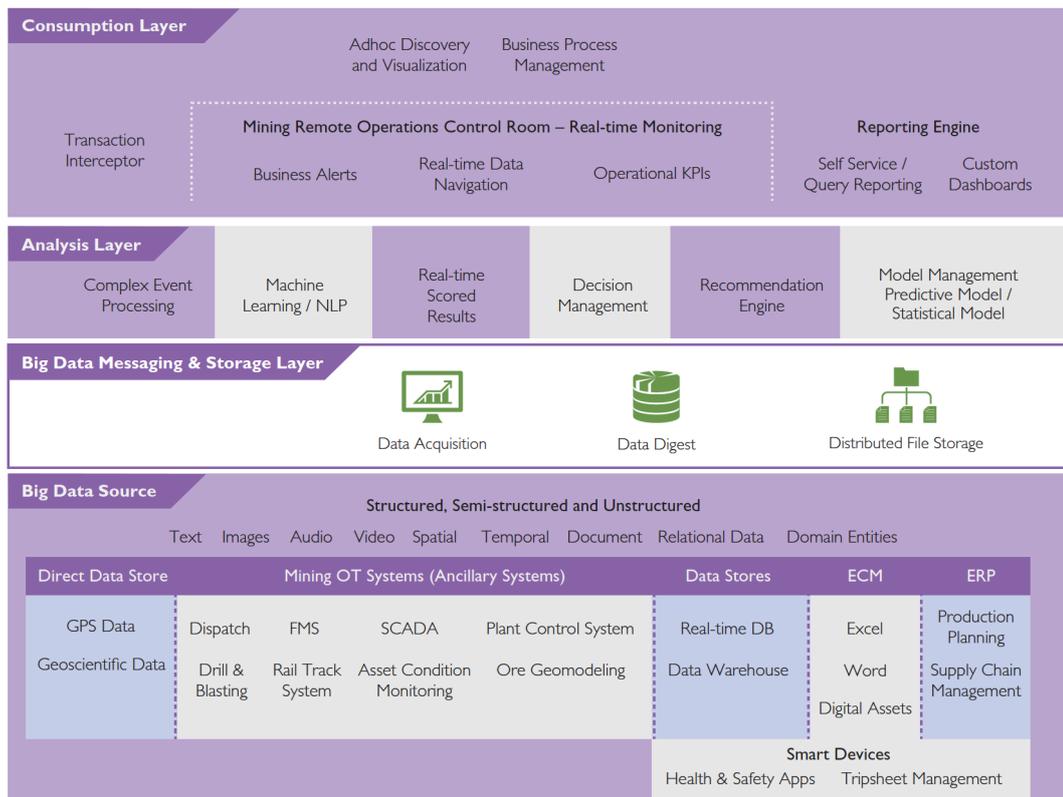


Figura 5: Painel estrutural de um sistema de Big Data Analytics
 Fonte: Adaptado de Chakraborti (2016)

1. *Big Data Source*: a primeira etapa deste sistema compreende nas fontes geradoras dos dados que podem ser estruturados, semiestruturados e não estruturados. Os dados são originados de diversos agentes, são eles: textos, imagens, áudios, espacial, temporal, documentos, dados relacionais e entidades de domínio (Chakraborti, 2016).

1.1. *Direct Data Store*: fontes de dados de medição direta, exemplos: informações de GPS e dados geocientíficos (Chakraborti, 2016).

1.2. *Mining Operational Technology Systems*: sistemas de tecnologia operacional para mineração, exemplos: *Dispatch*, *FMS*, *SCADA* e *Plan Control System*. Estes sistemas podem ser considerados como um sistema de gerenciamento de frotas, utilizam softwares especializados que utilizam informações de GPS e uma rede de rádio sem fio ou Wi-Fi. São capazes de rastrear e monitorar a produção, manutenção e segurança da mina. Alguns exemplos de aplicação: perfuração e dinamitação, sistema de trilhos, monitoramento de condição de ativos e geomodelagem do minério (DISPATCH®, 2017). A Figura 6 ilustra um típico sistema de gerenciamento de frotas.

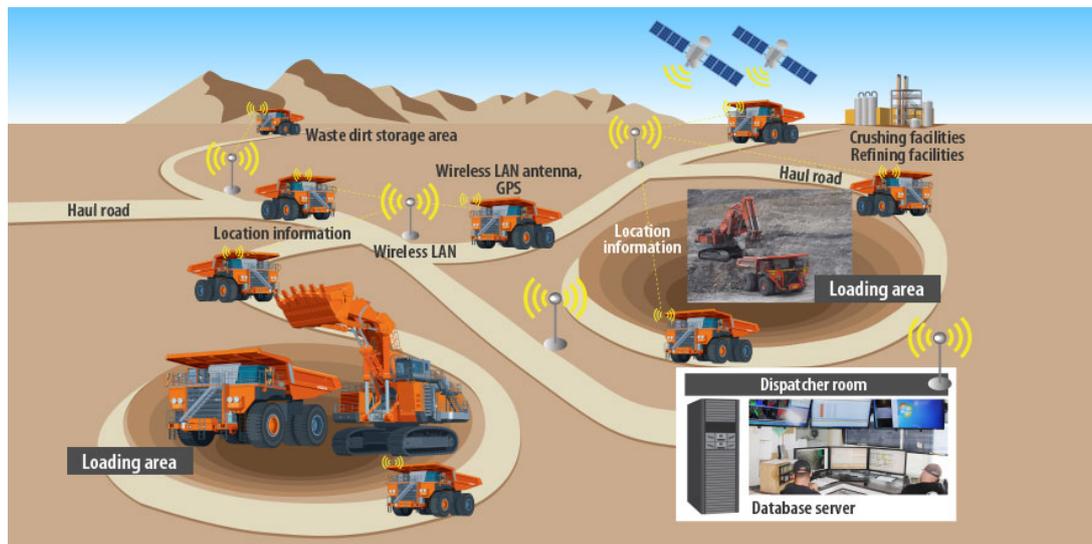


Figura 6: Sistema de gerenciamento de frotas

Fonte: Adaptado de Dispatch (2017)

1.3. *Data Stores*: banco de dados, exemplos: *Real-time Data Base* que significa base de dados em tempo real e *Data Warehouse* que é um depósito de dados digitais para armazenar informações detalhadas sobre o processo (Chakraborti, 2016).

1.4. *ECM (Enterprise Content Management)*: gestão de conteúdo empresarial é uma combinação de estratégias, métodos e ferramentas para captar, gerenciar, armazenar

e distribuir os conteúdos e documentos relacionados aos processos da mina. Exemplos: *Excel, Word, Digital Assets* (Ativos digitais) e dispositivos inteligentes como aplicativos de saúde e segurança e gerenciamento de planilhas (Chakraborti, 2016).

- 1.5. *ERP (Enterprise Resource Planning)*: planejamento de recurso empresarial é um sistema que integra todos os dados e processos de uma mina em um único sistema. Exemplos de aplicação: planejamento da produção e gestão da cadeia de abastecimento (Chakraborti, 2016).

2. *Big Data Messaging & Storage Layer*: a segunda etapa deste painel baseia-se no armazenamento dos dados e consiste de três fases. A primeira, *Data Acquisition* ou Aquisição de Dados trabalha com a obtenção e armazenamento dos dados gerados pelas fontes. A segunda fase chamada de *Digest Data* ou Digestão de Dados trabalha com a fragmentação dos dados, ou seja, divide os dados úteis daqueles que podem ser descartáveis. Por fim, a última etapa conhecida como *Distributed File Storage* ou Armazenamento de Arquivos distribuídos, dedica-se a organização e distribuição dos dados (Chakraborti, 2016).

3. *Analysis Layer*: o terceiro setor deste sistema de análise de grandes dados trabalha com o processamento e diagnóstico dos dados. Esta etapa é fundamental para que sejam selecionados aqueles dados de interesse e, posteriormente, a transformação destes em um formato a partir do qual as inferências possam ser obtidas. Chakraborti (2016) cita alguns exemplos de sistemas de análise de dados, são eles:
 - 3.1. *Complex Event Processing*: o processamento complexo de eventos pode ser resumido como o uso da tecnologia para prever eventos de alto nível, ou seja, refere-se a processos que coletam diferentes tipos de dados de diferentes partes de um sistema e procura resultados significativos para os tomadores de decisão.
 - 3.2. *Machine learning/Natural Language Processing*: a aprendizagem de máquina e o processamento de linguagem natural são sistemas que podem trabalhar juntos em problemas de fazer uma máquina “entender” a estrutura e o significado da linguagem natural usada pelos seres humanos.

- 3.3. *Real-time scored results*: resultados em tempo real são fundamentais para que seja rápida a tomada de decisões.
- 3.4. *Decision Management*: envolve todos os aspectos da concepção, construção e gerenciamento dos sistemas automatizados de tomada de decisão.
- 3.5. *Recommendation Engine*: motor de recomendação é um sistema de recomendação que funciona em fases lógicas bem definidas, que são coleta de dados, classificação e filtragem. Este sistema é capaz de reduzir, o que poderia se tornar uma decisão complexa, em apenas algumas recomendações de grande potencial.
- 3.6. *Model management / Predictive Model / Statistical Model*: gerenciamento de modelo, modelo preditivo e modelo estatístico são sistemas, que em conjunto, formam um processo de gerenciamento de modelos matemáticos e estatísticos para fazer previsões dos resultados.
4. *Consumption Layer*: a última etapa do sistema de análise de grandes dados é chamada de camada de consumo ou camada de saída de dados. Esta etapa fundamenta-se na transmissão aos interessados das informações recolhidas através da análise dos dados. Uma comunicação clara e concisa é fundamental para a correta interpretação dos resultados pelos tomadores de decisão. Estes resultados podem assumir a forma de relatórios, gráficos, tabelas e figuras (Marr, 2017). Para esta última etapa, Chakraborti (2016) cita a sala de controle remoto de uma mina (*Mining Remote Control Room*) que faz o monitoramento das atividades em tempo real. Como o objetivo deste trabalho é a atividade mineira, se faz necessário discursar sobre os três elementos presentes na sala de controle remoto que são apresentados por Chakraborti (2006), são eles:
- 4.1 *Business Alerts*: o primeiro elemento é chamado de alerta comercial e baseia-se no monitoramento dos dados, identificando atividades específicas e comunicando automaticamente que estas ações são apropriadas para os interessados.
- 4.2 *Real-time Data Navigation*: o segundo elemento é chamado de dados de navegação em tempo real e tem como objetivo informar sobre as atividades da mina imediatamente após a coleta dos dados, este processo não possui atrasos e é geralmente usado para navegações ou rastreamento.
- 4.3 *Operational KPIs (Operational Key Performance Indicators)*: o último elemento da sala de controle remoto é chamado de indicadores chave de desempenho operacional e

tem como meta fazer checagens regulares sobre a performance das atividades, determinando onde está ótimo e onde necessita de melhorias.

5.3 Casos de sucesso do uso de Big Data Analytics na mineração

5.3.1 A internet das coisas

O primeiro caso de sucesso apresentado por este trabalho trata-se de uma tecnologia que proporciona conectividade entre os operadores e os equipamentos da mina chamada de “A internet das coisas” ou *Internet of Things*. Esta tecnologia é a base para que uma empresa de mineração esteja totalmente conectada, na qual a empresa realiza plenamente a gestão dos equipamentos e o monitoramento da operação. Os dados são gerados constantemente a partir de sensores que medem gás e pressão de óleo, temperatura, peso e desempenho de cada equipamento em tempo real. De acordo com Mining Journal (2015), com a Internet da Coisas, a empresa de mineração possui visibilidade detalhada de suas operações e processos. Ao monitorar dispositivos de forma remota, os técnicos podem detectar equipamentos desgastados, como uma broca que está tornando-se maçante, e solicitar automaticamente um bit de reposição minimizando o tempo de inatividade do equipamento (Mining Journal, 2015).

O exemplo mais significativo do uso da Internet das Coisas na indústria mineral vem de minas que são operadas remotamente, usando caminhões sem motoristas e robótica. A Rio Tinto, uma respeitada empresa internacional de mineração, com produtos que abrangem recursos como o ferro, carvão, alumínio, ouro e urânio vem implementando esta tecnologia (Mining Journal, 2015). A mina de Pilbara na Austrália Ocidental é controlada a partir de um centro de operações remoto a 1500 km de distância em Perth. Os trens sem condutores carregam minério de ferro por 1500km de linha e a mina possui uma frota de caminhões inteligentes sem motoristas além de um controle remoto inteligente de perfuratrizes. A mina de Pilbara é um brilhante exemplo com o seu centro de operações remoto que com o auxílio da análise de grandes dados, gerou inteligência baseada em dados de equipamentos conectados, dispositivos e pessoas para otimizar processos e automatizar operações (Mining Journal, 2015).

5.3.2 Veículos inteligentes

O segundo caso de sucesso do uso da análise de grandes dados aborda o monitoramento da frota de equipamentos de uma mina. De acordo com o Mining Journal (2015), as companhias podem rastrear precisamente os veículos com sistemas que reconhecem se um caminhão está muito devagar alertando com sinais para a sala de controle. Estes sistemas ao seguir a rota de cada equipamento, sugerem ações para evitar acidentes que se, por ventura ocorrerem, eles identificam o local exato onde estão os veículos acidentados e quem os está operando. O uso da análise de grandes dados além de proporcionar mais segurança nas atividades da mina também proporciona mais eficiência destas atividades. Os sensores instalados nos equipamentos fornecem dados sobre o peso do caminhão, como foi carregado, sua condição, seu uso de combustível e a pressão dos pneus. Por exemplo, os vazamentos de combustível podem ser imediatamente detectados e os caminhões semi-carregados são carregados novamente antes de sair da frente de lavra (Mining Journal, 2015).

A Essar Steel, uma das principais produtoras de aço integrado da Índia com uma capacidade de produção anual de 10 milhões de toneladas, tem adotado a aplicação de dados importantes para o gerenciamento de seus veículos. A empresa tornou conectada toda a sua frota de veículos 3700 com GPS e RFID (Identificação por radiofrequência) para monitorar o movimento de cada equipamento, bem como checar níveis de combustível e uso. Os sensores foram colocados dentro de cada tanque de diesel para garantir a precisão da medição. A Essar Steel alcançou 5% de economia anual de manutenção e 10% de economia em combustível apenas usando esta tecnologia (Mining Journal, 2015).

5.3.3 Mídia Social e contribuição colaborativa (Crowdsourcing)

O terceiro e último exemplo de sucesso do uso de *Big Data Analytics* discorre sobre as mídias sociais e também da contribuição colaborativa, esta última consiste de um processo de obtenção de serviços e ideias através da solicitação de contribuições de um enorme grupo de pessoas, especialmente, de uma comunidade online. Estas duas formas de interação possuem papel importante em termos de reputação das empresas mineradoras e também no incentivo a participação na inovação e no financiamento (Mining Journal, 2015). Nos tempos atuais, a percepção pública das empresas de mineração é cada vez mais importante como identidade da

marca, podendo ter efeitos sobre a reputação corporativa, nos lucros e no valor de mercado. Assim, se faz necessário monitorar e analisar o sentimento público sobre a empresa e se atentar as questões relacionadas à empresa que estão sendo discutidas nas mídias sociais (Mining Journal, 2015).

De acordo com o *Mining Journal* (2015), esta análise do sentimento popular pode ajudar a prevenir ou atenuar manchetes negativas ou sentimentos negativos desenvolvidos pela comunidade local. Além disso, as empresas podem usar o monitoramento das redes sociais para se manterem informadas sobre os indivíduos de um determinado local e se estes estarão dispostos a trabalhar para a empresa e se a comunidade será receptiva à presença da mineradora na área de exploração.

Mining Journal (2015) cita o caso da empresa canadense Goldcorp que utilizou a contribuição colaborativa para ajudar a empresa a decidir o quê deveria ser feito com sua mina de 50 anos de exploração chamada de ' e que havia sido considerada como exaurida. Para tanto, a empresa abriu um concurso para obter uma resposta sobre os melhores lugares para perfurar dentro da sua área limite de exploração. Os prospectores virtuais de todo o mundo esperançosos em ganhar parte do prêmio oferecido pela Goldcorp no valor de US\$575.000, cruzaram dados disponibilizados pela empresa no seu próprio site e identificaram mais de 110 alvos propícios a conter ouro. No final, cerca de 80% daqueles alvos proporcionaram quantidades substanciais resultando em mais de US\$3 bilhões em ouro extraído (Mining Journal, 2015).

6 MINA DE OURO NA AUSTRÁLIA

O estudo foi realizado em uma planta de beneficiamento de uma mina de ouro na Austrália e as análises dos dados foram feitas usando dois softwares básicos, *Excel* e o *Matlab*. Para assegurar a confidencialidade dos dados, este trabalho não poderá citar o local da mina e a sua respectiva empresa.

6.1 Metodologia Analítica

A metodologia analítica usada no tratamento dos dados da planta de beneficiamento de ouro consiste de dois modelos estatísticos, são eles: a regressão linear e o *random forest*. Os dois modelos foram usados para analisar os dados de lixiviação de ouro (*LeachData*) e também os dados da perda de ouro (*GoldLoss*) gerando quatro modelos no total.

6.2 Estudo e análise dos dados

6.2.1 Dados de lixiviação da planta de beneficiamento de ouro (*LeachData*)

6.2.1.1 Regressão Linear

Para efeito didático, se faz necessário apresentar o banco de dados obtido da planta de beneficiamento de ouro. Na Tabela 3 são apresentadas as colunas com os valores de cada variável que será analisada.

Tabela 3: Dados de lixiviação da planta de beneficiamento de ouro

1	EXTR	RESGRAD	CN	AU	WESORE	AGIT	TEMP	SAKRES
2	94.56	89.56	185	3.272	13.70	17.35	25.0	20.4
3	94.22	82.46	214	3.583	15.24	18.36	26.6	37.0
4	93.94	91.50	135	3.200	11.01	34.23	26.0	42.2
5	95.36	89.11	189	3.165	13.45	17.47	27.8	89.2
6	91.57	80.18	247	3.797	13.71	12.00	23.3	10.6
7	94.21	84.20	295	3.507	15.24	49.64	26.2	38.5
8	95.39	81.18	158	3.667	13.96	12.38	23.3	55.5
9	91.17	81.98	226	3.591	13.71	12.86	21.8	0.0
10	95.20	86.55	153	3.539	9.88	11.29	22.9	81.9
11	93.82	81.29	163	3.285	14.44	15.36	20.0	19.2
12	94.64	85.01	205	3.302	13.74	20.07	26.8	39.0
13	96.63	88.88	176	3.883	13.96	9.97	24.0	64.6
14	94.93	90.73	175	3.530	10.99	22.27	24.3	40.9
15	93.85	85.06	171	4.132	13.70	17.55	26.4	17.8
16	95.38	89.38	158	3.854	8.44	15.19	25.9	89.1
17	95.90	90.00	145	3.858	10.30	16.86	26.7	129.6
18	95.33	92.13	154	3.662	11.00	20.14	26.9	41.7
19	95.62	90.44	179	3.541	11.41	10.28	21.0	120.8
20	92.40	80.77	156	3.593	13.70	12.55	22.5	10.2
21	95.84	91.33	239	3.125	12.36	14.89	28.4	92.0
22	94.53	82.41	170	3.436	18.34	15.28	20.0	113.7
23	95.23	83.55	159	3.458	13.97	13.97	23.8	58.9
24	94.25	87.23	151	3.306	11.42	13.79	20.5	62.3
25	94.05	80.36	149	3.313	16.43	14.38	20.1	57.6
26	94.97	91.40	151	3.937	10.99	21.29	25.8	41.7
27	94.63	85.50	222	3.634	13.76	22.58	27.5	69.7
28	95.11	86.46	217	3.536	13.45	18.87	27.4	84.7
29	91.91	80.07	233	2.833	13.71	10.54	23.6	14.6
30	94.19	82.87	173	3.720	13.97	21.56	25.5	58.1
31	94.83	90.31	190	3.521	11.37	22.67	24.5	41.0
32	93.82	84.60	308	3.254	15.25	55.95	26.3	38.2
33	95.48	85.27	157	3.959	13.97	15.55	26.2	80.1
34	96.14	89.26	149	4.349	8.45	16.44	27.5	86.8
35	95.96	88.94	170	3.739	10.70	16.93	28.0	99.9
36	94.33	90.33	189	3.667	24.05	7.96	19.1	135.6
37	95.48	90.34	175	3.274	12.34	15.84	28.1	92.3
38	96.80	91.30	178	3.651	10.70	12.33	27.4	92.0
39	95.31	86.00	154	3.711	11.55	13.03	24.3	81.0
40	93.74	83.24	182	3.434	11.86	9.12	21.4	59.0
41	90.91	79.55	288	3.709	13.71	12.99	22.5	0.0
42	93.88	83.20	254	3.368	15.24	33.62	26.2	37.8
43	94.72	91.78	169	3.766	20.88	10.44	21.2	137.0
44	92.71	79.32	153	3.637	13.70	12.53	23.1	18.0
45	94.57	89.83	168	3.555	11.41	10.75	18.9	83.2
46	93.09	83.07	180	3.414	12.69	9.25	22.0	48.7
47	94.37	76.20	199	4.210	15.23	16.86	27.1	69.0
48	96.01	91.42	222	3.182	10.69	15.44	27.3	92.0
49	94.26	75.66	201	4.129	15.23	16.15	27.1	64.0
50	95.87	86.80	170	4.018	8.44	12.86	25.1	82.4
51	95.13	90.74	158	3.448	11.00	19.15	27.3	40.0
52	92.36	81.50	200	3.677	12.71	9.31	23.4	19.6
53	95.57	90.00	152	3.661	8.45	16.30	27.6	79.9
54	93.93	88.13	138	3.293	11.42	20.03	21.3	61.9
55	94.44	88.43	188	3.688	23.61	10.11	19.3	126.1

O papel que cada variável exerce na planta de beneficiamento é mostrado na Tabela 4.

Tabela 4: Representatividade de cada variável dos dados de lixiviação na planta de beneficiamento de ouro

EXTR	Porcentagem de extração de ouro	ENTRADA
RESGRAD	Concentração residual de ouro	ENTRADA
CN	Concentração de cianeto	ENTRADA
AU	Ouro	ENTRADA
WESORE	Composição do minério	ENTRADA
AGIT	Agitação	ENTRADA
TEMP	Temperatura	ENTRADA
SAKRES	Consumo de reagente	SAÍDA

É possível notar que há sete variáveis de entrada (independentes) e apenas uma variável de saída (dependente) que é considerada o objetivo do modelo, chamada aqui de SAKRES e

representante do consumo de reagente na planta. A regressão linear foi desenvolvida no programa *Excel* e consistiu das três partes seguintes:

7.2.1.1.1 Primeiramente, foi necessário normalizar todos os dados de entrada para construir o modelo completo que representa a relação entre os sete dados de entrada e o único dado de saída. A Figura 7 mostra a relação entre os dados de saída medidos do consumo de reagente e os dados de saída previstos para o consumo de reagente calculado pelo modelo completo. O valor de R^2 mostrado no gráfico é chamado de coeficiente de determinação e indica o quanto o modelo é capaz de explicar a variável de saída em relação as variáveis de entrada.

7.2.1.1.2 Na segunda parte foi formulado mais sete modelos de acordo com a ausência de uma das variáveis de entrada. O processo consiste em excluir a primeira variável de entrada (EXTR) e gerar o primeiro modelo, então, retorna-se com esta variável para a sua coluna de origem e exclui-se a próxima variável de entrada (RESGRAD,) gerando o segundo modelo e assim sucessivamente. No final do processo é possível comparar o quanto cada variável é significativa para o modelo completo analisando o decréscimo do coeficiente de determinação..

7.2.1.1.3 Por fim, construiu-se o gráfico de barras mostrado na Figura 8 para checar a importância que cada variável de entrada exerce para o modelo em comparação com as outras variáveis de entrada. Este gráfico foi criado baseado nos valores absolutos dos coeficientes fornecidos pelo modelo completo da regressão linear.

O modelo completo da regressão linear para os dados de lixiviação encontra-se anexado ao Apêndice B deste trabalho.

6.2.1.2 *Random Forest*

O modelo não-linear *random forest* usado para a análise dos dados de lixiviação foi operado no programa estatístico *Matlab* e dividido em três partes como no caso da regressão linear.

7.2.1.2.1 A primeira parte baseia-se na normalização dos dados conforme o mesmo tratamento usado para a regressão linear. Depois disso, os dados são apresentados e rodados no Matlab, os comandos usados para esta análise estão presentes no Apêndice C deste trabalho. No gráfico da Figura 9 é possível constatar a relação entre o consumo de reagente medido e o consumo de reagente previsto fornecido pelo modelo completo. O valor de R^2 mostrado na Figura 9 significa o quanto é explicativo o modelo completo do random forest.

7.2.1.2.2 A segunda parte é similar à da regressão linear (ver seção 7.2.1.1.2), onde há a exclusão de cada variável de entrada, entretanto, antes de rodar os dados no Matlab, é necessário que cada dado de entrada se torne aleatório em um processo chamado de randomize, assim são construídos mais sete modelos de random forest.

7.2.1.2.3 Na última parte foi construído um gráfico de barras mostrado na Figura 10 que representa a diferença entre o valor de R^2 do modelo completo e o valor do R^2 de cada modelo random forest gerado com cada dado aleatório. Esta diferença significa a importância que cada variável de entrada possui para o modelo.

6.2.2 Dados da perda de ouro (Goldloss data)

6.2.2.1 Regressão linear

A análise dos dados pelo modelo de regressão linear da perda de ouro é a mesma usada para os dados de lixiviação. A única diferença está no tamanho dos dados da perda de ouro com 14 variáveis de entrada e 2 variáveis de saída. Por se tratar de uma tabela com muitos dados, estas informações foram anexadas no Apêndice A deste trabalho.

Na Tabela 5 é mostrado o quê cada variável presente nos dados da perda de ouro representa na planta de beneficiamento.

Tabela 5: Representatividade das variáveis dos dados da perda de ouro na planta de beneficiamento.

HG	Teor principal	ENTRADA
RG	Teor de resíduo	ENTRADA
ST	Tonelagem de solução	ENTRADA
TT	Tonelagem tratada	ENTRADA
FF	Filtro de alimentação	ENTRADA
FW	Filtro de lavagem	ENTRADA
SD	Filtro Solids duty	ENTRADA
FA	Filtro ARLA	ENTRADA
SND	Solução duty	ENTRADA
ES	Entrada de solução	ENTRADA
FFL	Filtro Floc	ENTRADA
FV	Filtro a vácuo	ENTRADA
NACN1	Agitador NaCn 1	ENTRADA
NACN2	Agitador NaCn 2	ENTRADA
UG	Perdas de ouro não dissolvidas	SAÍDA
DG	Perdas de ouro dissolvidas	SAÍDA

Nota-se que neste caso existem duas variáveis de saída (UG e DG), o objetivo deste trabalho será considerar apenas as perdas de ouro não dissolvidas. Como o tratamento dos dados já foi explicado anteriormente na seção 7.2.1.1, este trabalho irá focar apenas em descrever os passos tomados neste processo, são eles:

7.2.2.1.1 Foi construído o gráfico mostrado na Figura 11 que relaciona os dados das perdas de ouro medidos e os dados das perdas de ouro previstos pelo modelo completo da regressão linear. O modelo completo da regressão linear para os dados da perda de ouro encontra-se no Apêndice D.

7.2.2.1.2 Foi criado 14 modelos de previsão de acordo com a ausência de cada variável.

7.2.2.1.3 O gráfico mostrado na Figura 12 representa a importância de cada variável de entrada para o modelo completo da regressão linear.

6.2.2.2 Random Forest

O tratamento dos dados pelo modelo *random forest* para a perda de ouro é o mesmo tratamento usado para os dados de lixiviação. Os seguintes passos foram seguidos para a análise dos dados.

7.2.2.2.1 Foi construído o gráfico mostrado na Figura 13 que apresenta a relação entre os dados medidos das perdas de ouro não dissolvidas e os dados previstos das perdas de ouro não dissolvidas.

7.2.2.2.2 Foi desenvolvido 14 modelos random forest como descrito para os dados de lixiviação.

7.2.2.2.3 O gráfico da Figura 14 apresenta a relevância de cada variável de entrada para o modelo random forest em comparação com as outras variáveis.

6.3 Resultados

6.3.1 Dados de lixiviação

6.3.1.1 Regressão linear

O valor do coeficiente de determinação (R^2) mostrado no gráfico da Figura 7 representa o quanto os dados do modelo estão próximos da linha de regressão ajustada, em outras palavras, representa o quanto o modelo pode explicar o comportamento da variável SAKRES. Este valor varia entre 0,0 e 1,0 e quanto mais próximo de 1,0 mais o modelo pode ser considerado explicativo.

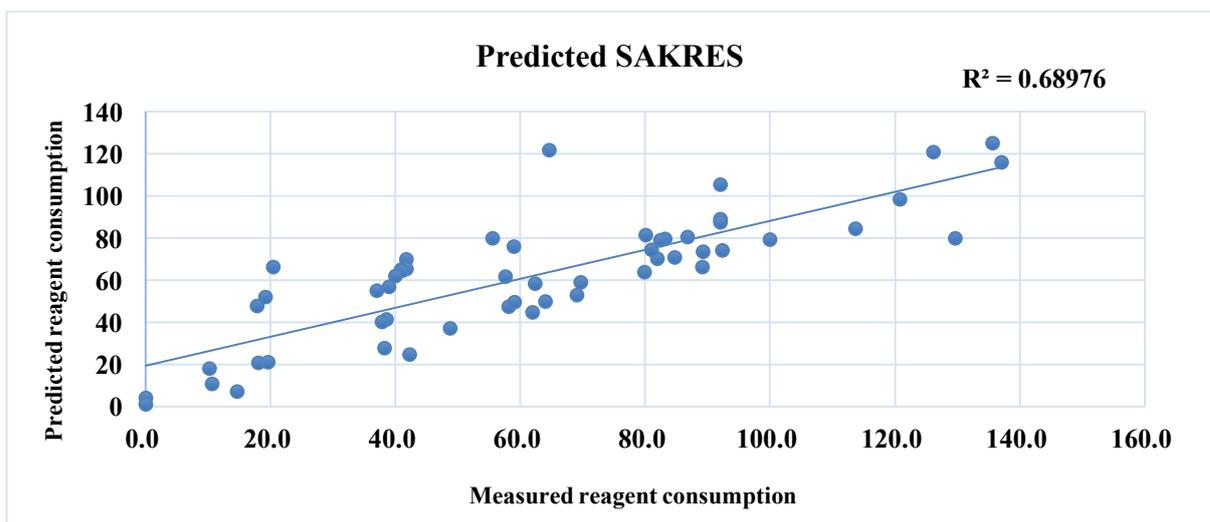


Figura 7: Consumo de reagente medido x Consumo de reagente previsto – Dados de lixiviação

Neste caso, o valor de $R^2 = 0,68976$ significa que o modelo de regressão linear pode ser considerado incompetente, pois o valor de 0,68976 é relativamente distante de 1,0. A Tabela 6

mostra os valores do coeficiente de determinação para os sete modelos de previsão que foram construídos com a ausência de suas respectivas variáveis.

Tabela 6: Valores de R^2 para os modelos de previsão conforme a ausência de cada variável – Dados da lixiviação

	Variável excluída do modelo	Valor de R^2
Modelo 1	EXTR	0,46645
Modelo 2	RESGRAD	0,67057
Modelo 3	CN	0,68213
Modelo 4	AU	0,68461
Modelo 5	WESORE	0,62455
Modelo 6	AGIT	0,66299
Modelo 7	TEMP	0,67366

Estes resultados indicam que a variável EXTR, que representa a porcentagem de extração do ouro, possui maior significância para o modelo que as outras variáveis. Isto se deve ao fato do seu $R^2 = 0,46645$ ter sido o valor que mais diferiu em relação ao R^2 do modelo completo que foi de 0,68976. Assim, a ausência da variável EXTR gera a maior deficiência para a explicabilidade do modelo indicando que aproximadamente apenas 47% da variabilidade da variável dependente (SAKRES) consegue ser explicada pelas outras variáveis presentes no modelo. A segunda mais importante variável para o modelo é a WESORE, que representa a composição do minério, com valor de $R^2 = 0.62455$. Observe a Figura 8.

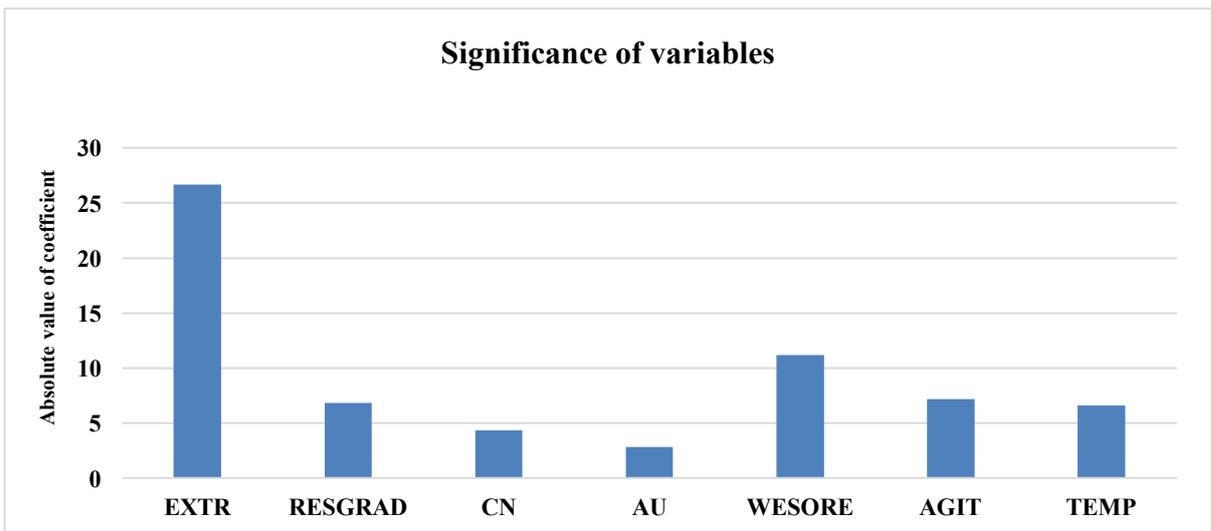


Figura 8: Valores absolutos dos coeficientes de cada variável de entrada – Regressão linear - Dados de lixiviação

Observando o gráfico de barras da Figura 8 é possível verificar quantitativamente a influência de cada variável de entrada através dos valores absolutos dos coeficientes fornecidos pelo modelo completo da regressão linear. Portanto, confirma-se aqui que as variáveis EXTR e WESORE exercem de fato um significativo papel na regressão linear para os dados de lixiviação.

6.3.1.2 *Random Forest*

Geralmente, uma regressão não linear como o modelo *random forest* apresenta resultados mais precisos do que a regressão linear. Entretanto, no caso dos dados da lixiviação, isto não se comprova. O valor de R^2 do modelo completo de *random forest* foi menor do que o apresentado pela regressão linear implicando em um modelo ruim e preciso. Observe o gráfico mostrado pela Figura 9.

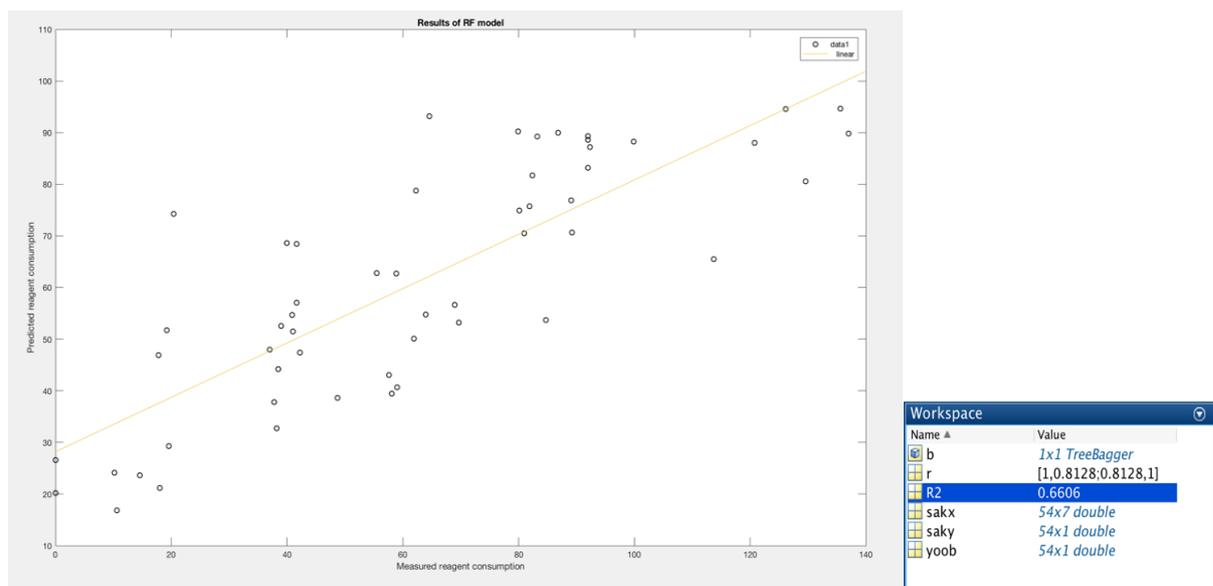


Figura 9: Consumo de reagente medido x Consumo de reagente previsto – Dados da lixiviação

Observa-se que o valor de $R^2 = 0,6606$ continua sendo muito abaixo daquilo que é considerado satisfatório. pode ser considerado relativamente próximo de 1,0 e, assim, apresentar um certo grau de confiabilidade.

A tabela 7 mostra os valores de R_{X_i} obtidos através dos sete modelos *random forest* desenvolvidos conforme a ausência de cada variável. Estes valores serão usados para

demonstrar a importância que cada variável exerce para o modelo *random forest*. Esta importância é mensurada através do valor absoluto das diferenças entre R^2 e R_{xi} , onde o valor de R^2 é o valor do modelo completo (0,6606). O gráfico de barras da Figura 10 apresenta estas relevâncias.

Tabela 7: Valores de R_{xi} para os modelos *random forest* conforme a ausência de cada variável – Dados de lixiviação

	Variável excluída do modelo	Valor de R_{xi}
Modelo 1	EXTR	0.5254
Modelo 2	RESGRAD	0.6875
Modelo 3	CN	0.6349
Modelo 4	AU	0.6658
Modelo 5	WESORE	0.6171
Modelo 6	AGIT	0.5984
Modelo 7	TEMP	0.6063

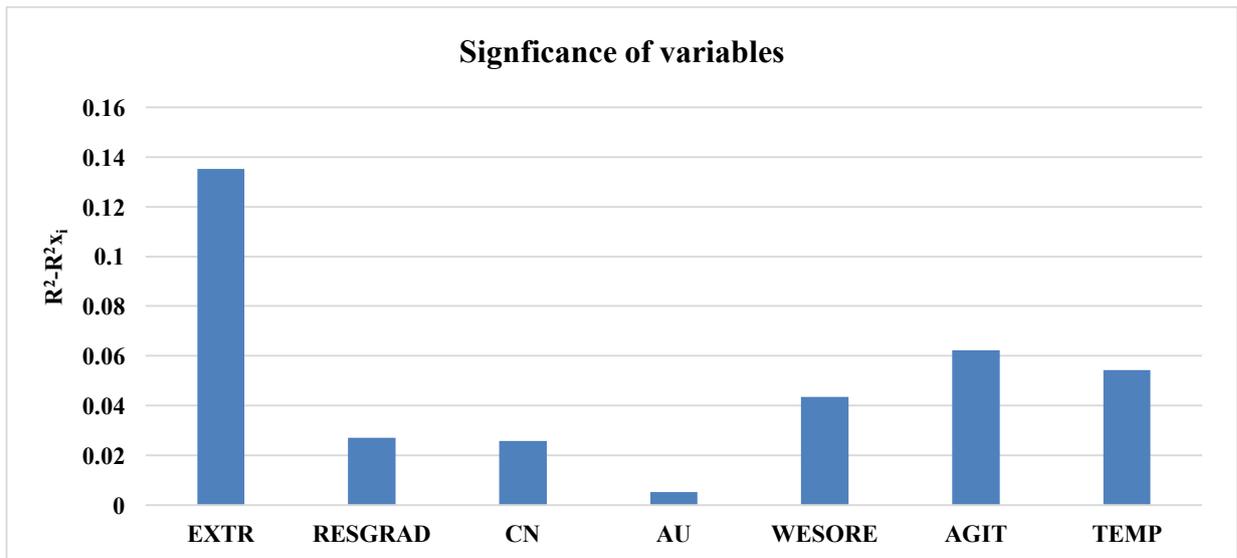


Figura 10: Importância das variáveis através da diferença entre R^2 e R_{xi} – Dados da lixiviação

É possível verificar que a porcentagem de extração de ouro (EXTR) possui a maior importância para o modelo *random forest* assim como na regressão linear. Entretanto, diferentemente da regressão linear, a variável de agitação (AGIT) apresenta a segunda maior significância para o modelo em questão.

6.3.2 Dados da perda de ouro

6.3.2.1 Regressão linear

O valor do coeficiente de determinação apresentado pelo modelo da regressão linear para os dados de perda de ouro foi de 0,74719. Isto resulta em um modelo razoável e melhor do que aquele desenvolvido para os dados de lixiviação. Observe o gráfico da Figura 11.

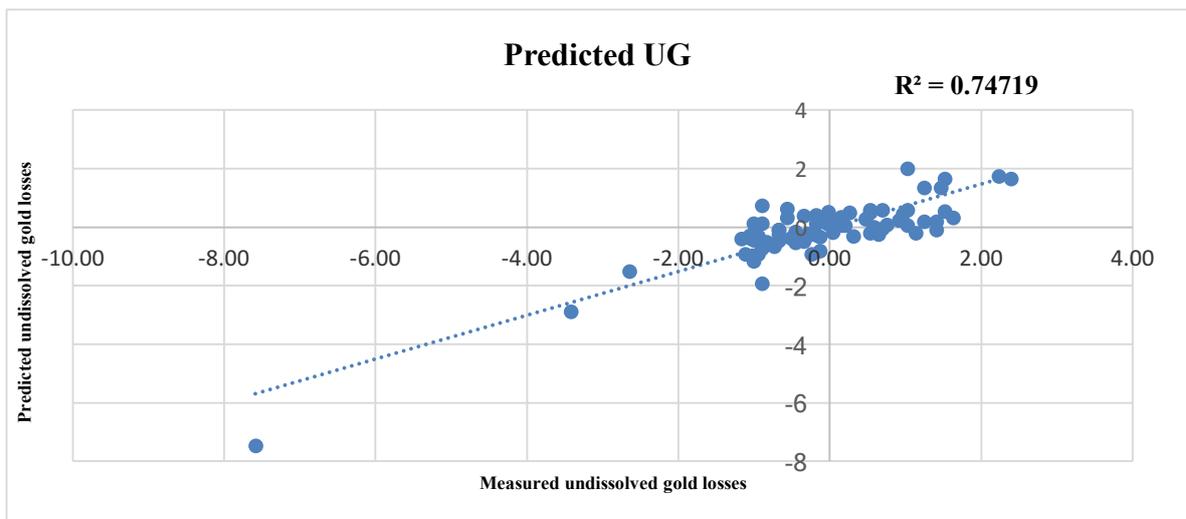


Figura 11: Perda de ouro não-dissolvido medido x Perda de ouro não-dissolvido previsto

Na tabela 8 são apresentados os coeficientes de determinação dos modelos preditivos conforme a ausência das respectivas variáveis.

Tabela 8: Valores de R^2 para os modelos de regressão linear conforme a ausência de cada variável – Dados da perda de ouro

	Variável excluída do modelo	Valor de R^2
Modelo 1	HG	0,7126
Modelo 2	RG	0,7370
Modelo 3	ST	0,7468
Modelo 4	TT	0,7233
Modelo 5	FF	0,7438
Modelo 6	FW	0,7394
Modelo 7	SD	0,7235
Modelo 8	FA	0,7193
Modelo 9	SND	0,7468
Modelo 10	ES	0,7467
Modelo 11	FFL	0,7470
Modelo 12	FV	0,7379
Modelo 13	NACN1	0,6676
Modelo 14	NACN2	0,6921

No gráfico de barras apresentado na Figura 12 estão os valores absolutos dos coeficientes fornecidos pelo modelo completo da regressão linear.

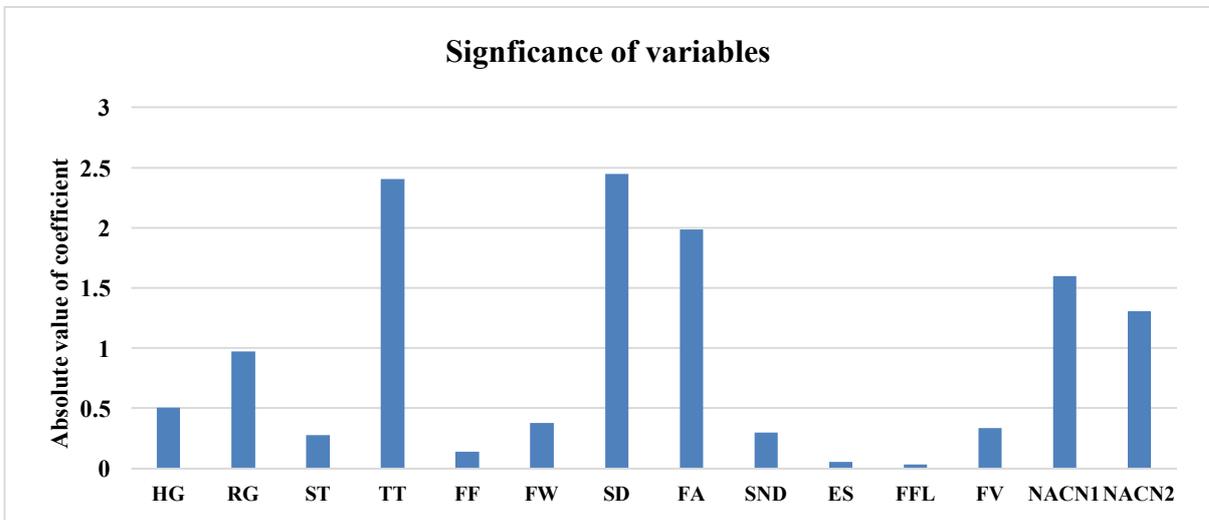


Figura 12: Valores absolutos dos coeficientes de cada variável de entrada – Dados da perda de ouro – Regressão linear

É possível observar que as variáveis que representam, respectivamente, o filtro *solids duty* e a tonelagem tratada são as variáveis que exercem o papel mais significativo no modelo em questão.

6.3.2.2 *Random forest*

Como citado anteriormente, comumente, os modelos não lineares como o *random forest* apresentam análises mais confiáveis e precisas. Entretanto, para os dados da perda de ouro, isto não foi comprovado. O coeficiente de determinação para este modelo foi de 0,3274, muito abaixo dos valores considerados satisfatórios ou próximos de 1,0. Assim, para este caso, o modelo pode ser considerado ineficiente ou não explicativo, pois apenas 37% da variável dependente (UG) pode ser explicada pelo modelo. Observe o gráfico da Figura 13.

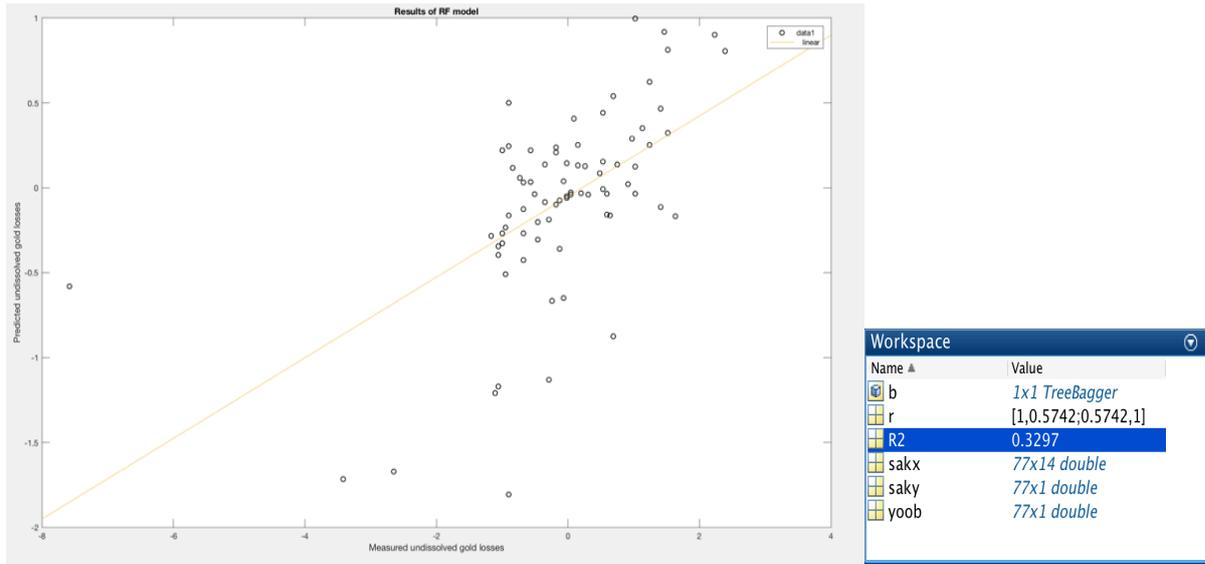


Figura 13: Consumo de reagente medido x Consumo de reagente previsto – Perda de ouro – *Random Forest*

Foi estimada a influência exercida por cada variável no modelo *random forest* para os dados da perda de ouro da seguinte maneira: calculou-se os valores absolutos da diferença entre o coeficiente de determinação do modelo completo ($R^2=0,3297$) e dos coeficientes de determinação (R_{X_i}) obtidos em cada modelo onde houve a exclusão de uma variável de entrada. Os valores de R_{X_i} são mostrados na tabela 9. Com os valores desta diferença, foi construído o gráfico mostrado na Figura 14.

Tabela 9: Valores de R_{X_i} para os modelos *random forest* conforme a ausência de cada variável – Dados da perda de ouro

	Variável excluída do modelo	Valores de R_{X_i}
Modelo 1	HG	0.2616
Modelo 2	RG	0.3250
Modelo 3	ST	0.2948
Modelo 4	TT	0.2371
Modelo 5	FF	0.3284
Modelo 6	FW	0.2972
Modelo 7	SD	0.3066
Modelo 8	FA	0.3048
Modelo 9	SND	0.2255
Modelo 10	ES	0.2488
Modelo 11	FFL	0.3010
Modelo 12	FV	0.2968
Modelo 13	NACN1	0.3392
Modelo 14	NACN2	0.3317

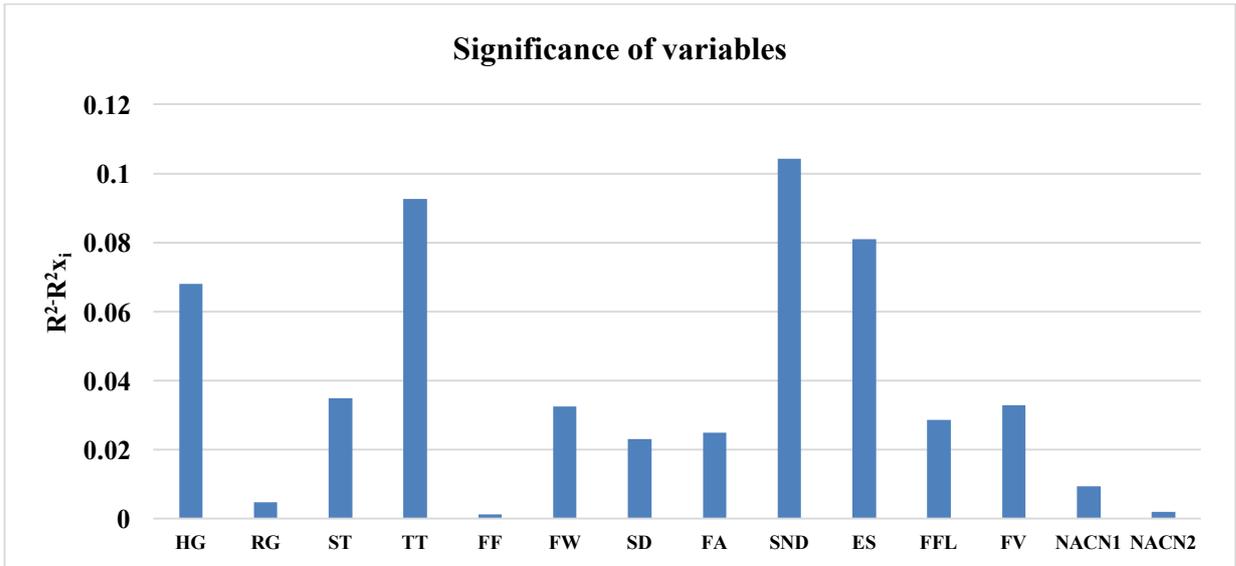


Figura 14: Importância das variáveis através da diferença entre R^2 e $R^2_{X_i}$ - Perda de ouro – *Random Forest*

Verifica-se, portanto, que a variável que representa a solução *duty* (SND) exerce o papel mais importante para o modelo *random forest* desenvolvido para os dados da perda de ouro. Isto difere do resultado do modelo linear para os mesmos dados, onde a variável que representa o filtro *solids duty* (SD) representou a maior importância. Nos dois modelos, a variável da tonelagem tratada (TT) apresentou a segunda maior importância.

7 CONCLUSÃO

Diante da relevância exercida pela indústria mineral nos dias atuais, se faz necessário o progresso desta atividade em detrimento da manutenção do padrão de vida da sociedade moderna. Assim, visando um progresso sustentável, as companhias mineradoras têm considerado aplicar novas tecnologias como o uso das análises de grandes dados (*Big Data Analytics*). O uso destes grandes dados pode ser considerado como uma reformulação de uma tecnologia já existente, ou seja, é a análise de dados gerados por equipamentos autônomos durante uma atividade específica. Devido aos inúmeros benefícios que este novo conhecimento pode trazer para um projeto, como o quanto de água pode ser economizada no processamento mineral, esta ciência tem ganhado espaço na indústria mineral, considerada como fator transformador da mineração nos próximos cinco a dez anos.

Para exemplificar o uso da análise destes grandes dados, o presente estudo apresentou análises de dois bancos de dados oriundos de uma planta de beneficiamento de uma mina de ouro na Austrália. Para esta análise, foram utilizados dois modelos preditivos, a regressão linear e o *random forest* gerando. Se forem corretamente estudados e aplicados, a regressão linear e o *random forest* podem otimizar qualquer processo dentro de uma mina. Geralmente, o *random forest* é mais preciso e consegue analisar os grandes dados rapidamente. Entretanto, isto não foi confirmado por este trabalho, nos dois bancos de dados, a regressão linear apresentou mais confiabilidade devido o seu valor do coeficiente de determinação. O modelo de regressão linear para os dados da perda de ouro foi o modelo que apresentou a melhor explicabilidade dentre os quatro modelos desenvolvidos por este estudo. Como os quatro modelos preditivos apresentaram resultados aquém do satisfatório, funções matemáticas não lineares podem explicar melhor o comportamento dos dados estudados. Funções como a polinomial e a exponencial, talvez resultariam em resultados mais confiáveis e explicativos acerca do comportamento das variáveis.

Finalmente, o incentivo a pesquisa e desenvolvimento de novas técnicas de engenharia é de suma importância para o desenvolvimento de uma sociedade cada vez mais interessada com os aspectos ambientais e sociais. Portanto, objetivando estes interesses, a análise de grandes dados poderá ser a chave para o sucesso dos empreendimentos minerários, tornando, em um futuro próximo, a indústria mineral em uma indústria de tecnologia da informação.

REFERÊNCIAS BIBLIOGRÁFICAS

- AIIM. (2017). *What is Enterprise Content Management*. Disponível em: <http://www.aiim.org/What-is-ECM-Enterprise-Content-Management#>. Acesso em junho de 2017.
- Alpaydm, E. (2010). *Introduction to Machine Learning* (2nd Edition ed.). Cambridge, MA, USA: The MIT Press.
- Auret, L. & Aldrich, C. (2012). *Interpretation of nonlinear relationships between process variables by use of random forests*. Minerals Engineering 35, 27-42.
- Big data business. (2017). *Você sabe o que é Big Data Analytics*. Disponível em: <http://www.bigdatabusiness.com.br/voce-sabe-o-que-e-big-data-analytics/>. Acesso em junho de 2017.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.
- Chakraborti, S. (2016). *Application of Big Data Solution to Mining Analytics*. Disponível em: <http://www.wipro.com/documents/application-of-big-data-solution-to-mining-analytics.pdf>. Acesso em junho de 2017.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). *Anomaly detection: A survey*. *Journal ACM Computing Surveys*, 41 (3).
- Conway, J. (2014). *General Principles – Linear Regression*. Disponível em: <https://alpine.atlassian.net/wiki/pages/viewpage.action?pageId=8683787>. Acesso em junho de 2017.
- Farah, B. N. (2016). *Big data - what data and why?* *Journal of Management Policy and Practice*, 17(1), 11-17. Disponível em: <http://search.proquest.com.dbgw.lis.curtin.edu.au/docview/1788513651?accountid=10382>. Acesso em junho de 2017.
- General Kinematics. (2017). *A brief history of mining and the advancement of mining technology*. Disponível em: <https://www.generalkinematics.com/blog/a-brief-history-of-mining-and-the-advancement-of-mining-technology/>. Acesso em junho de 2017.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA, USA: The MIT Press.
- Harrison, Thomas H. *Intranet data warehouse*. São Paulo: Berkeley Brasil, 1988.
- Hunt, D. (2014). *Big Data Challenges: Volume, Variety, Velocity & Veracity*. NC State University Research (Fall), 2-8.

IBM. (2012). *What is big data?* Disponível em: <http://www-01.ibm.com/software/in/data/bigdata/>. Acesso em junho de 2017.

IBM. (2013, August 22). *The Four V's of Big Data*. Disponível em: <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>>. Acesso em junho de 2017.

IBM. (2013, July 30). *The FOUR V's of Big Data*. Disponível em: http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-bigdata.jpg>. Acesso em junho de 2017.

IBRAM. (2017). Disponível em: http://www.ibram.org.br/150/15001002.asp?ttCD_CHAVE=166525. Acesso em junho de 2017.

IBRAM. (2017). Disponível em: <http://www.ibram.org.br/sites/1300/1382/00005836.pdf>. Acesso em junho de 2017.

Jensen, S. (2013). *Electric Underground*. Disponível em: <http://www.oemoffhighway.com/electronics/article/11224086/electrification-of-underground-mining-equipment>>. Acesso em junho de 2017.

Marr, B. (2017). *Big Data: The 4 Layers Everyone Must Know*. Disponível em: <https://www.ap-institute.com/big-data-articles/big-data-the-4-layers-everyone-must-know>.> Acesso em junho de 2017.

Mining Journal (2015). *Mining Big Data Guide*. Disponível em: http://europe.prod.cdn.pressrun.com/aspermontmedia/miningjournal/webapp_preview/index.html#issue/miningbigdataguide2015/portrait/1.>. Acesso em junho de 2017.

National Research Council (U.S.). (2002). *Evolutionary and revolutionary technologies for mining*. Washington, D.C: National Academy Press.

Nguyen, J. (2017). *Regression Basics For Business Analysis*. Disponível em: <http://www.investopedia.com/articles/financial-theory/09/regression-analysis-basics-business.asp>>. Acesso em junho de 2017.

Ohlhorst, F. J. (2012). *Wiley and SAS Business Series : Big Data Analytics : Turning Big Data into Big Money*. Somerset, NJ, USA: John Wiley & Sons.

OMA. (2017). *100 innovations in the mining industry*. https://www.oma.on.ca/en/ontariomining/resources/Minalliance_100_innovaations_en.pdf. Acesso em junho de 2017.

Refermat, E. (2015). *What is big data? Automatic Merchandiser*, 57, 22-26. Disponível em: <http://search.proquest.com.dbgw.lis.curtin.edu.au/docview/1685178997?accountid=10382>>. Acesso em junho de 2017.

Reffat, R. M., Gero, J. S., & Peng, W. (2004). *Using data mining on building maintenance during the building life cycle*. 38th Australian & New Zealand Architectural Science Association (pp. 91-97). Hobart: University of Tasmania.

Russom, P. (2011). *Big data analytics*. TDWI best practices report, fourth quarter, 19, 40.

SAS. (2017). *What is Big Data*. https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html. Acesso em junho de 2017.

Significados. (2017). *Data Warehouse*. <https://www.significados.com.br/data-warehouse/>. Acesso em junho de 2017.

Surange, S., & Bansal, A. (2013). *IBM Accelerator for Machine Data Analytics, Part 1: Speeding up machine data analysis*. Disponível em: <<http://www.ibm.com/developerworks/data/library/techarticle/dm-1301machinedata1/>>. Acesso em junho de 2017.

Suthaharan, S. (2015). *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Integrated Series Information. Springer 36, 359.

Technopedia. (2017). *Complex event processing (CEP)* <https://www.techopedia.com/definition/29496/complex-event-processing-cep>. Acesso em junho de 2017.

Wencomine. (2017). *The most common fleet management system fms question*. <http://www.wencomine.com/the-most-common-fleet-management-system-fms-question/>. Acesso em junho de 2017.

Wills, B.A. (2013). *Mineral Processing Technology: An Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery*. International Series on Materials Science and Technology. Elsevier, 4, 804.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, X., Zhu, X., Wu, G., & Ding, W. (2014). *Data mining with big data*. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107. Disponível em: <<http://dx.doi.org/dbgw.lis.curtin.edu.au/10.1109/TKDE.2013.109>>. Acesso em junho de 2017.

APÊNDICE A – DADOS DA PERDA DE OURO

	HG	RG	ST	TT	FF	FW	SD	FA	SND	ES	FFL	FV	NACN1	NACN2	UG	DG
1																
2	0.29	-1.03	-0.51	-2.86	0.02	1.21	-2.12	-1.70	-0.67	-0.07	-2.03	0.69	-0.21	-0.48	-1.00	-0.43
3	0.05	0.56	0.95	0.24	-1.08	0.84	0.74	-1.24	1.59	-0.66	-0.82	0.78	-1.41	-1.52	-0.57	0.07
4	-0.14	-0.75	-0.28	-0.04	-0.52	0.05	-0.10	-0.30	-0.14	0.22	-0.19	0.70	0.48	0.41	-0.73	1.39
5	-0.06	1.95	0.59	0.19	0.30	0.05	-0.24	1.02	0.15	0.54	0.11	0.61	-2.55	-1.88	1.24	-1.09
6	-2.45	1.11	-1.72	-5.36	2.15	6.34	-5.05	1.45	-2.20	-3.88	6.39	-1.67	1.07	1.36	-3.42	-0.10
7	1.16	-0.19	0.50	0.17	0.22	0.03	-0.21	0.89	0.12	1.15	0.20	0.69	-0.21	-0.48	-0.07	0.07
8	-2.98	-1.03	1.53	0.41	-0.09	0.70	0.60	0.42	1.64	0.87	-0.22	0.65	0.82	0.94	-0.07	-1.09
9	-1.29	0.93	-0.19	0.53	-0.84	-0.21	0.70	-0.76	0.13	0.03	-0.22	0.50	-0.51	-0.80	-0.95	0.73
10	-1.22	-1.40	-2.08	0.14	-0.58	-1.49	0.28	-0.44	-1.97	-1.02	-0.30	0.70	0.48	0.41	-1.11	-0.10
11	0.30	-0.84	-0.27	0.38	0.17	-0.65	0.23	0.23	-0.37	0.18	0.93	0.74	0.86	1.01	1.02	-0.92
12	1.81	0.28	1.14	0.69	-0.98	0.59	0.49	0.28	1.02	1.19	-0.92	0.65	0.82	0.94	-0.90	-0.43
13	0.38	-1.12	-2.79	0.22	-0.55	-1.56	0.30	-0.30	-2.03	-0.40	0.18	-1.50	-0.46	-0.75	-0.29	1.56
14	-0.73	2.05	-0.71	-0.49	0.41	-0.47	-0.71	0.66	-0.97	-1.00	0.27	-1.67	1.07	1.36	-0.95	-0.59
15	0.41	0.65	-0.72	0.33	-0.93	-0.47	0.18	0.26	-0.83	-0.07	0.13	0.59	0.08	-0.13	-1.00	2.38
16	0.04	-0.84	-0.46	-0.13	-0.49	-0.17	0.11	-0.61	-0.21	0.27	-0.39	-1.48	1.18	1.56	0.97	0.23
17	-1.01	-0.10	1.83	1.44	-0.81	0.89	0.51	0.54	1.58	-0.51	-1.27	0.78	-1.41	-1.52	-0.13	-1.91
18	0.84	-0.84	-0.31	1.02	-0.09	-0.94	0.65	0.59	-0.55	1.43	-0.54	-1.71	0.33	0.21	-0.02	-0.26
19	-0.79	-0.10	0.66	1.07	-0.34	-0.36	0.87	0.17	0.39	0.24	-0.69	-1.65	0.21	0.05	0.15	-0.43
20	-0.10	1.77	-0.90	-0.40	0.37	-0.67	-0.55	0.08	-1.09	-0.57	0.41	-1.67	1.07	1.36	-1.17	0.73
21	-0.38	0.46	1.07	0.66	-0.99	0.57	1.10	-1.16	1.68	0.82	-1.07	-1.65	0.21	0.05	1.63	0.56
22	-0.16	-0.19	1.66	0.35	6.19	0.66	-0.36	1.76	0.85	-0.45	0.85	0.70	0.48	0.41	-0.46	0.90
23	-0.22	-1.12	-0.22	0.70	-0.04	-0.72	0.98	-0.82	0.13	-0.83	0.33	0.35	-0.89	-1.15	0.59	0.40
24	-0.52	-0.19	-1.63	0.39	-0.28	-1.43	2.04	-3.42	-0.24	-0.50	0.61	0.74	0.86	1.01	-0.02	-0.43
25	0.82	1.02	-0.20	-1.13	0.69	0.35	-1.03	-0.05	-0.19	1.51	0.55	0.61	-2.55	-1.88	1.46	0.90
26	-2.22	0.46	-0.81	-2.68	2.17	0.93	-2.58	0.52	-1.02	-3.19	1.73	0.63	-0.02	-0.26	-0.90	-1.58
27	0.61	1.67	1.09	0.15	0.18	0.48	-0.22	0.89	0.69	0.26	-0.13	0.61	-2.55	-1.88	2.39	0.73
28	-0.63	-0.75	-0.76	1.00	-0.53	-1.05	0.82	0.14	-0.85	-0.38	0.08	-1.50	-0.46	-0.75	0.04	0.07
29	0.24	0.83	-0.01	-0.23	0.38	0.12	1.14	-3.02	1.51	0.91	-0.18	0.59	0.08	-0.13	-0.68	-0.76
30	1.59	-1.22	0.85	-0.67	0.59	0.83	-0.90	0.81	0.49	0.91	0.15	0.38	0.57	0.56	1.52	0.90
31	-0.44	0.65	-1.09	0.35	-0.15	-1.08	1.17	-1.92	-0.32	-0.05	-0.53	-1.48	1.18	1.56	0.48	-0.26
32	1.28	0.28	1.18	0.34	0.51	0.26	0.72	-1.00	1.71	-0.21	-1.17	0.65	0.82	0.94	-0.57	-1.25
33	0.02	0.37	0.00	-0.19	-0.53	0.17	-0.52	0.90	-0.37	1.03	0.16	0.69	-0.21	-0.48	-1.06	-0.76
34	0.19	0.83	0.34	0.23	0.11	-0.09	-0.19	0.98	-0.07	0.13	0.01	0.69	-0.21	-0.48	-1.06	-0.59
35	-0.76	0.83	-0.16	0.58	-0.75	-0.32	1.03	-1.18	0.36	-0.33	-0.90	0.50	-0.51	-0.80	-0.35	0.56
36	0.10	0.09	0.45	-0.37	0.19	0.42	-0.66	0.85	0.09	0.42	-0.48	0.59	0.08	-0.13	-0.68	0.07
37	-1.53	-0.19	0.14	0.22	-0.32	-0.04	0.70	-1.22	0.71	-0.53	0.51	0.35	-0.89	-1.15	-0.84	0.40
38	0.03	-0.10	0.41	0.48	0.02	-0.17	0.37	0.10	0.37	-0.38	-0.23	-1.48	1.18	1.56	0.53	-0.43
39	1.35	1.95	0.88	0.60	0.10	0.04	0.00	1.29	0.32	0.68	-0.43	0.77	0.57	0.55	1.24	-0.59
40	0.00	-1.12	0.96	0.05	0.00	0.54	-0.38	1.08	0.48	0.21	-0.27	0.38	0.57	0.56	0.09	-0.10
41	0.62	1.11	-0.17	-0.04	-0.17	-0.16	0.14	-0.50	0.04	0.85	-0.58	0.77	0.57	0.55	1.13	0.23
42	-2.52	0.00	-2.13	-0.82	0.14	-1.28	-1.27	1.56	-2.61	-2.09	1.35	0.63	-0.02	-0.26	-2.65	1.56
43	0.59	-0.75	-1.06	1.03	-0.88	-1.11	1.12	-0.46	-0.90	0.59	0.22	-1.50	-0.46	-0.75	0.91	-0.43
44	0.28	-0.28	1.66	0.71	-0.71	0.75	0.82	-0.47	1.80	-0.72	-0.46	0.65	0.82	0.94	-0.18	-1.09
45	0.41	-0.47	0.79	-0.06	1.75	-0.20	-0.31	0.65	0.32	-0.48	-0.23	0.38	0.57	0.56	-0.18	-1.58
46	-0.15	-0.75	-0.30	-0.81	0.34	0.12	-0.50	-0.67	-0.02	-0.86	-0.16	0.35	-0.89	-1.15	-0.51	1.06
47	1.25	-0.56	0.82	-1.53	0.65	1.42	-1.43	0.07	0.49	-2.25	-0.51	0.38	0.57	0.56	0.69	-2.08
48	1.03	-3.74	0.13	-0.05	0.58	-0.23	-0.03	-0.09	0.17	-0.19	1.31	0.35	-0.89	-1.15	1.02	0.56
49	-1.29	1.02	1.63	0.47	-0.97	1.10	0.32	0.21	1.54	-0.64	-0.71	0.65	0.82	0.94	-0.13	-1.42
50	-1.46	0.09	0.39	1.14	-0.81	-0.25	1.50	-1.04	0.89	-0.43	-0.12	-1.65	0.21	0.05	1.41	-0.10
51	-0.99	0.18	-1.04	0.20	-0.40	-0.85	0.30	-0.33	-0.93	-0.92	0.02	-1.67	1.07	1.36	-0.46	-0.59
52	0.57	-1.12	0.42	-0.04	0.00	0.06	0.39	-0.62	0.71	0.20	-0.13	0.38	0.57	0.56	0.69	1.23
53	1.94	-0.28	0.13	0.29	0.80	-0.56	0.39	-0.34	0.28	0.60	-0.18	0.69	-0.21	-0.48	-0.35	0.73
54	0.56	-0.94	0.44	-0.01	0.03	0.34	-0.59	1.53	-0.01	0.51	0.09	0.77	0.57	0.55	1.41	-0.10
55	0.71	-0.94	-0.31	1.10	-0.31	-0.89	0.65	0.76	-0.61	1.74	-0.11	-1.71	0.33	0.21	0.26	0.73
56	0.79	-0.38	1.02	1.06	0.12	-0.18	0.28	1.70	0.32	1.26	-0.69	0.77	0.57	0.55	0.53	-0.10
57	0.81	0.93	-1.81	-2.09	1.08	-0.36	-1.84	-0.22	-1.76	-1.24	1.07	-1.67	1.07	1.36	-1.06	-1.09
58	0.36	1.02	-0.34	0.20	-0.98	-0.11	0.09	-0.20	-0.26	1.06	0.85	0.59	0.08	-0.13	-0.90	-0.26
59	0.86	-0.94	-0.86	0.02	0.72	-1.06	0.26	-0.65	-0.61	0.20	1.79	0.74	0.86	1.01	0.75	0.23
60	-0.28	0.46	0.84	-0.15	0.05	0.59	-0.42	0.71	0.53	0.18	0.15	0.77	0.57	0.55	-0.18	1.56
61	0.40	1.67	0.98	0.56	-0.09	0.22	-0.14	1.58	0.29	0.72	-0.57	0.61	-2.55	-1.88	1.52	-1.58
62	1.24	0.56	0.57	-0.60	0.42	0.49	-0.55	-0.02	0.44	0.75	0.13	0.61	-2.55	-1.88	2.23	-0.59
63	0.56	-0.19	-0.80	0.07	-0.61	-0.52	0.02	0.07	-0.84	0.55	-0.11	-1.50	-0.46	-0.75	0.20	-0.26
64	0.00	0.93	-0.87	-1.39	-0.78	0.63	-1.13	-0.45	-0.71	-0.36	-1.41	0.50	-0.51	-0.80	-0.24	1.72
65	0.41	-1.31	-0.76	0.11	-0.16	-0.69	0.07	0.03	-0.78	0.56	0.39	0.63	-0.02	-0.26	-0.29	-0.26
66	0.63	0.93	0.97	0.27	1.20	-0.12	-0.35	1.52	0.31	0.26	-0.33	0.70	0.48	0.41	-1.00	-1.42
67	-0.36	0.56	1.38	0.62	-1.43	1.00	1.15	-1.35	2.13	-0.01	-0.27	0.78	-1.41	-1.52	-0.02	-0.76
68	0.66	-1.78	-0.92	-0.13	-0.36	-0.57	-0.05	-0.22	-0.86	1.81	-0.22	-1.71	0.33	0.21	0.15	-0.76
69	-1.25	0.09	-1.13	-0.64	-0.77	-0.22	-0.45	-0.39	-1.00	-0.26	-0.22	-1.48	1.18	1.56	0.31	0.07
70	-1.45	0.18	1.65	1.15	-0.86	0.58	1.33	-0.66	2.03	-0.48	-1.02	-1.65	0.21	0.05	-0.90	2.55
71	-0.64	0.65	-0.20	-0.59	-0.28	0.29	-0.54	-0.06	-0.19	-0.50	-0.43	-1.48	1.18	1.56	0.59	0.23
72	-0.37	-1.40	-0.38	-0.02	-0.28	-0.28	0.06	-0.24	-0.29	0.02	0.78	0.35	-0.89	-1.15	0.53	0.90
73	0.16	0.00	-0.94	0.21	0.01	-0.96	0.14	0.08	-0.98	-0.32	0.68	0.74	0.86	1.01	0.64	-0.26
74	1.66	0.56	0.49	0.17	-0.44	0.29	-0.10	0.60	0.23	0.10	-0.13	0.70	0.48	0.41	0.04	2.38
75	0.01	-0.38	-0.90	0.50	-0.71	-0.79	0.56	0.10	-0.95	-0.80	-0.09	0.50	-0.51	-0.80	-0.68	1.39
76	-0.46	-0.10	-0.45	0.34	-0.12	-0.62	-0.31	1.58	-1.04	0.73	0.62	0.63	-0.02	-0.26	-0.68	-0.10
77	0.82	-0.19	0.65	0.35	0.39	-0.06	0.53	-0.55	0.92	1.90	-0.19	0.61	-2.55	-1.88	1.02	-0.92
78	-11.30	-71.49	-7.99	-8.25	-7.21	-4.34	-7.48	-19.01	-8.10	-6.95	-2.81	-16.43	-2.68	-1.89	-7.58	-2.90

APÊNDICE B – MODELO COMPLETO DA REGRESSÃO LINEAR PARA OS DADOS DE LIXIVIAÇÃO

SUMMARY OUTPUT											
Regression Statistics											
Multiple R	0.83051912										
R Square	0.689762008										
Adjusted R Square	0.642551879										
Standard Error	21.11711504										
Observations	54										
ANOVA											
		df	SS	MS	F	Significance F					
Regression		7	45606.9776	6515.28252	14.6104664	7.72895E-10					
Residual		46	20512.8972	445.932548							
Total		53	66119.8748								
		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%		
Intercept		62.35185185	2.87367537	21.6975976	8.4111E-26	56.56744334	68.1362604	56.5674433	68.1362604		
EXTR	26.6871888	26.68718878	4.63787346	5.75418649	6.754E-07	17.3516337	36.0227438	17.3516337	36.0227438		
RESGRAD	6.85100402	6.851004018	4.06133213	1.68688593	0.09839655	-1.324033547	15.0260416	-1.3240335	15.0260416		
CN	4.35704775	4.357047746	4.09647969	1.0636078	0.29305812	-3.888738203	12.6028337	-3.8887382	12.6028337		
AU	2.82025976	2.820259763	3.22628861	0.87414987	0.38657771	-3.67392238	9.31444191	-3.6739224	9.31444191		
WESORE	11.164621	11.16462103	3.59057379	3.10942531	0.00321327	3.937170848	18.3920712	3.93717085	18.3920712		
AGIT	7.1734912	-7.173491204	3.60080084	-1.9921933	0.05230439	-14.42152736	0.07454495	-14.421527	0.07454495		
TEMP	6.62692991	-6.626929907	4.28831255	-1.5453468	0.12911507	-15.25885536	2.00499555	-15.258855	2.00499555		
RESIDUAL OUTPUT											
Observation	Predicted SAKRES	Residuals	Observation	Predicted SAKRES	Residuals	Observation	Predicted SAKRES	Residuals	Observation	Predicted SAKRES	Residuals
1	66.34347973	-45.94348	15	66.46498705	22.6350129	29	47.65045904	10.449541	43	20.9604584	-2.9604584
2	55.31435222	-18.314352	16	79.97510236	49.6248976	30	64.60394196	-23.603942	44	79.68001637	3.51998363
3	24.78891531	17.4110847	17	70.05174998	-28.35175	31	27.81818456	10.3818154	45	37.16386124	11.5361388
4	73.76984657	15.4301534	18	98.57671988	22.2232801	32	81.54183946	-1.4418395	46	52.98469483	16.0153052
5	10.85963009	-0.2596301	19	18.19422287	-7.9942229	33	80.77059849	6.02940151	47	87.58835813	4.41164187
6	41.51216937	-3.0121694	20	89.01624264	2.98375736	34	79.4295013	20.4704987	48	49.91786153	14.0821385
7	80.15002525	-24.650025	21	84.56877938	29.1312206	35	125.376058	10.223942	49	79.20132562	3.19867438
8	4.059827763	-4.0598278	22	76.21110538	-17.311105	36	74.32237226	17.9776277	50	62.04136877	-22.041369
9	70.25969068	11.6401093	23	58.55366813	3.74633187	37	105.5314336	-13.531434	51	21.168431	-1.568431
10	52.22627186	-33.026272	24	61.80012333	-4.2001233	38	74.54384398	6.45615602	52	63.87448613	16.0255139
11	57.02815975	-18.02816	25	65.49898026	-23.79989	39	49.7651454	9.2348546	53	44.8355839	17.0644061
12	121.8597001	-57.2597	26	58.98864616	10.7113538	40	1.162334236	-1.1623342	54	120.9822322	5.11776783
13	65.19430569	-24.294306	27	70.990641	13.709359	41	40.39008852	-2.5900885			
14	47.96635531	-30.166355	28	7.361852734	7.23814727	42	116.1088502	20.8911498			

APÊNDICE C – COMANDOS USADOS PARA O MODELO RANDOM FOREST NO MATLAB

Getting data into Matlab workspace

```
>> sakx = []; %Copy input data from Excel file and paste
between the square brackets;

>> saky = []; %Copy target data from Excel file and paste
between the square brackets;
```

OR

```
>> load c:\folder\sakres.txt %After saving Excel data as *.txt
file in c:\folder

>> sakx = sakres(:,1:end-1); %input consists of all records of
columns one to the second last one

>> saky = sakres(:,end); %target consists of all records of
the last column
```

Fitting the model

```
>> b =
TreeBagger(200,sakx,saky,'method','regress','oobpred','on');
%Model fitted with 200 trees, method tells model to do
regression, and oobpred tells it to automatically assess model
against unseen data |

>> yoob = oobPredict(b); %Generates model outputs on unseen
data ('oob' = out-of-bag' data)
```

Plotting the results

```
>> figure; plot(saky,yoob,'ko');

>> xlabel('Measured reagent consumption');

>> ylabel('Predicted reagent consumption');

>> title('Results of RF model');
```

Calculating model fit (R² value)

```
>> r = corrcoef(saky,yoob);

>> R2 = r(1,2)*r(1,2); %Fraction of variance of target
variable explained by the model;
```

APÊNDICE D – MODELO COMPLETO DA REGRESSÃO LINEAR PARA OS DADOS DA PERDA DE OURO

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.8644								
R Square	0.74718736								
Adjusted R Square	0.69010063								
Standard Error	0.73299059								
Observations	77								
ANOVA									
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression		14	98.4507937	7.03219955	13.0886358	1.3238E-13			
Residual		62	33.3110631	0.53727521					
Total		76	131.761857						
Coefficients									
		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept		-0.0984864	0.08353203	-1.1790252	0.24289299	-0.2654645	0.06849178	-0.2654645	0.06849178
HG	0.5024133	0.5024133	0.17253538	2.91194369	0.00498629	0.15752	0.8473066	0.15752	0.8473066
RG	0.97004591	-0.9700459	0.61298741	-1.5824891	0.11862625	-2.1953903	0.25529846	-2.1953903	0.25529846
ST	0.27444432	0.27444432	0.93588011	0.29324731	0.77031245	-1.5963534	2.14524201	-1.5963534	2.14524201
TT	2.40293293	-2.4029329	0.99300958	-2.4198487	0.01847216	-4.3879308	-0.4179351	-4.3879308	-0.4179351
FF	0.1369632	-0.1369632	0.14945453	-0.9164206	0.36299727	-0.4357185	0.16179214	-0.4357185	0.16179214
FW	0.37622591	-0.3762259	0.27304968	-1.3778662	0.17319763	-0.9220444	0.16959262	-0.9220444	0.16959262
SD	2.44885448	2.44885448	1.01665493	2.40873714	0.01899308	0.41659024	4.48111873	0.41659024	4.48111873
FA	1.98686718	1.98686718	0.75942324	2.61628441	0.01115097	0.46880176	3.50493261	0.46880176	3.50493261
SND	0.2961235	0.2961235	0.95598643	0.30975701	0.75778448	-1.6148661	2.20711314	-1.6148661	2.20711314
ES	0.05439351	0.05439351	0.16254585	0.33463485	0.73903	-0.270531	0.37931803	-0.270531	0.37931803
FFL	0.03042875	-0.0304287	0.15775662	-0.1928841	0.84768018	-0.3457797	0.28492224	-0.3457797	0.28492224
FV	0.33502893	-0.3350289	0.22240406	-1.5063975	0.13704352	-0.7796083	0.10955045	-0.7796083	0.10955045
NACN1	1.5971092	-1.5971092	0.36142057	-4.4189771	4.0611E-05	-2.3195786	-0.8746398	-2.3195786	-0.8746398
NACN2	1.30749987	1.30749987	0.35582304	3.67457903	0.00049898	0.59621975	2.01877999	0.59621975	2.01877999

RESIDUAL OUTPUT									
<i>Observation</i>	<i>Predicted UG</i>	<i>Residuals</i>	<i>Observation</i>	<i>Predicted UG</i>	<i>Residuals</i>	<i>Observation</i>	<i>Predicted UG</i>	<i>Residuals</i>	<i>Observation</i>
1	-1.1702636	0.16543652	36	-0.5107534	-0.3296081	71	-0.2109478	0.74113291	
2	0.32568213	-0.8919343	37	0.5878562	-0.0576711	72	-0.2529205	0.89274939	
3	-0.6721601	-0.0585577	38	0.19448233	1.04838701	73	0.25754478	-0.2207564	
4	1.35251237	-0.109643	39	0.14722425	-0.055614	74	-0.0985467	-0.5773492	
5	-2.8991886	-0.5178005	40	-0.2070346	1.34026018	75	-0.463972	-0.2119238	
6	0.1723767	-0.2452321	41	-1.5227469	-1.126736	76	1.99130147	-0.9677196	
7	0.28156218	-0.3544175	42	0.22544929	0.68848887	77	-7.4800621	-0.1033886	
8	-0.9217767	-0.0282285	43	0.41209233	-0.5945914				
9	-0.9250015	-0.1894693	44	0.11364704	-0.2961461				
10	0.05075825	0.97282364	45	-0.3777081	-0.1337222				
11	0.74046246	-1.6356458	46	-0.0504393	0.74509005				
12	-0.2915869	-0.0005559	47	0.58422205	0.43935985				
13	-0.3217968	-0.6282084	48	-0.3331597	0.20548246				
14	-0.4455463	-0.5592808	49	-0.0991384	1.50647336				
15	0.44404036	0.52471967	50	-0.1590517	-0.2975567				
16	-0.8146632	0.68698595	51	0.58198406	0.11266665				
17	0.52154642	-0.5395799	52	0.37788768	-0.7248524				
18	0.05847413	0.08795796	53	0.1852318	1.22210313				
19	-0.40748	-0.7618126	54	0.48574621	-0.2296704				
20	0.31231335	1.31430904	55	0.4878527	0.04233242				
21	-0.529526	0.07291761	56	-0.2820402	-0.7776087				
22	0.00193351	0.58307347	57	-0.7183075	-0.1768757				
23	0.09832901	-0.1163625	58	0.08000321	0.66946937				
24	1.33403706	0.12811973	59	-0.2742382	0.09173906				
25	-1.9388482	1.04366485	60	1.65834159	-0.1413629				
26	1.65649269	0.73763578	61	1.7455543	0.48410859				
27	-0.1744638	0.21125218	62	0.0594145	0.14183945				
28	-0.0968275	-0.5790684	63	-0.9155629	0.67824197				
29	0.5372159	0.97976276	64	-0.2177642	-0.0743786				
30	0.28544005	0.18992321	65	0.11700838	-1.1218354				
31	0.62118344	-1.1874356	66	0.42322006	-0.4412536				
32	-0.3693346	-0.6903143	67	0.33201951	-0.1855874				
33	-0.3047477	-0.7549013	68	-0.3202537	0.63115139				
34	-0.4863716	0.13940694	69	0.11335712	-1.0085404				
35	-0.25915	-0.4167458	70	-0.0158706	0.60087757				