



Escola de Tecnologias e Arquitetura

Departamento de Ciências e Tecnologias da Informação

Fatores de sucesso na avaliação de questões de escolha múltipla: o caso de exames de Excel.

Daniela Patrícia de Barros Almeida

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em
Informática e Gestão

Orientador:

Dr. Sérgio Moro, Professor Auxiliar ISCTE, Instituto Universitário de Lisboa

Coorientador:

Dr. António Martins, Professor Auxiliar ISCTE, Instituto Universitário de Lisboa

Setembro, 2017

Resumo

A avaliação é uma componente de ensino-aprendizagem, realizada continuamente na sala de aula, pelo professor e pelo aluno e constitui um importante dispositivo pedagógico para o sucesso escolar dos alunos. No processo de avaliação, cabe ao professor procurar meios de melhorar a sua tarefa, de modo a construir métodos de avaliar, que vão de encontro ao que é lecionado nas aulas. Esta tarefa não é fácil e requer que o docente progrida continuamente, acompanhando a constante evolução dos avanços tecnológicos e educativos. Deste modo, pretende-se com o desenvolvimento deste projeto procurar fatores importantes do estudo que irão futuramente auxiliar os docentes na construção de métodos de avaliação.

Esta dissertação foca-se na análise de perguntas de escolha múltipla de exames de Excel realizados no ISCTE-IUL (Instituto Universitário de Lisboa), com vista a encontrar quais os fatores que influenciam a classificação obtida pelo aluno nos exames. Pretende-se, a partir desta análise, extrair conhecimento que possibilite a percepção da forma que cada um dos fatores analisados afetam os resultados, e que venha a contribuir para melhorar o desempenho do docente na construção de avaliações, tanto na estrutura de elaboração das questões, como no conteúdo a ser abordado em cada uma.

Com este intuito, os dados da amostra serão utilizados na aplicação de técnicas de *data mining* para optimização das avaliações. Desta investigação irá resultar um modelo explicativo ilustrando quais das variáveis consideradas são relevantes para o estudo e qual a importância de cada uma.

Deste modo, é possível extrair conhecimento útil que poderá suportar as decisões dos docentes na construção das avaliações futuras. O estudo destina-se a docentes que procuram melhorar o seu desempenho na elaboração de questões de escolha única.

Abstract

Assessment is a teaching-learning component, carried out continuously in the classroom, by the teacher and by the student and is an important pedagogical device for the students' academic success. In the evaluation process, it is up to the teacher to find ways to improve his / her task, in order to construct methods of evaluation, which are in agreement with what is taught in class. This task is not easy and requires that the teacher progresses continuously, following the constant evolution of technological and educational advances. In this way, it is intended with the development of this project to look for important factors of the study that will in the future assist the teachers in the construction of evaluation methods.

This dissertation focuses on the analysis of multiple choice questions from Excel exams at ISCTE-IUL (Instituto Universitário de Lisboa), in order to find out which factors influence the student's classification in the exams. The aim of this analysis is to extract knowledge that allows the perception of the way each of the analyzed factors affect the results, and that will contribute to improve the performance of the teacher in the construction of evaluations, both in the structure of elaboration of the questions, as in the content to be addressed in each one, ensuring that the knowledge to be evaluated goes against what is taught in class.

For this purpose, the sample data will be used in the application of data mining techniques to optimize the evaluations. From this investigation will result an explanatory model illustrating which of the variables considered are relevant to the study and the importance of each.

In this way, it is possible to extract useful knowledge that can support the decisions of the teachers in the construction of the future evaluations. The study is intended for teachers seeking to improve their performance in elaborating single-choice questions.

Índice

Resumo	i
Abstract	ii
Índice de Figuras	iv
Índice de Tabelas	v
Lista de Abreviaturas	vi
1. Introdução	1
1.1. Enquadramento e Problema	1
1.3. Estrutura da dissertação	2
2. Revisão da literatura	4
2.1. Avaliação académica.....	4
2.2. Taxonomia de Bloom.....	4
2.3. Questões de escolha múltipla	7
2.4. Avaliação de desempenho académico em cursos de MS Excel	9
2.5. Casos de estudo.....	10
2.5.1. Análise de dados de avaliação académica.....	10
2.5.2. Metodologias usadas na construção dos testes dos casos revistos	12
2.6. Data mining e extração de conhecimento	12
2.6.1. Métodos de abordagem em Data mining	15
2.6.2. Técnicas de Modelação em Data mining	18
2.7. Data mining na avaliação académica	22
3. Metodologia de investigação	24
3.1. Trabalho realizado	25
4. Resultados e Discussão	36
4.1. Análise das métricas e Variáveis	36
4.2. Conclusões	57
4.3. Limitações e Trabalho Futuro	59
BIBLIOGRAFIA	60
ANEXO	64

Índice de Figuras

Figura 1 - Processo de KDD, adaptado de Han e Kamber (2006).....	13
Figura 2 - Fases da metodologia CRISP-DM, adaptado de Chapman et al. (2000)	15
Figura 3 - Processo de Data Mining: SEMMA, adaptado de Turban et al. (2011)	17
Figura 4 - Taxonomia do Data Mining, adaptado de Maimon & Rokach (2010)	19
Figura 5 - Processo de modelação	33
Figura 6 – Processo de Extração de Conhecimento	38
Figura 7 - Curvas ROC, MLPE e SVM.....	40
Figura 8 - Curvas ALIFT, MLPE e SVM.....	40
Figura 9 - Relevância das variáveis.....	42
Figura 10 – Número de Palavras, "Certo"	43
Figura 11 – Número de Palavras, "Não respondido"	44
Figura 12 - Bloco de Matéria, "Certo".....	44
Figura 13 - Bloco de Matéria, "Não respondido"	45
Figura 14 - Grau de Dificuldade, "Certo"	46
Figura 15 - Grau de Dificuldade, "Não respondido"	47
Figura 16 - Número de Próximas, "Certo".....	47
Figura 17 - Número de Próximas, "Não respondido"	48
Figura 18 - Notas de EM, "Certo"	49
Figura 19 - Notas de EM, "Não respondido"	49
Figura 20 - Número de caracteres, "Certo"	50
Figura 21 - Número de caracteres, "Não respondido"	51
Figura 22 - Áreas de Curso, "Certo".....	51
Figura 23 - Áreas de Curso, "Não respondido".....	52
Figura 24 - Número de Testes, "Certo"	53
Figura 25 - Número de Testes, "Não respondido"	54
Figura 26 - Unidade Curricular, "Certo"	54
Figura 27 - Unidade Curricular, "Não respondido".....	55

Índice de Tabelas

Tabela 1 - Análise das variáveis.....	30
Tabela 2 -AUC e ALIFT Preditivos	37
Tabela 3 - Análise das métricas AUC e ALIFT	39
Tabela 4 - Taxonomia Original, Fonte: Bloom et al.(1956), Bloom (1986), Driscoll (2000) e Krathwohl (2002).	65
Tabela 5 - Taxonomia Revisto, Fonte: Anderson, Krathwohl e Airasian (2001).....	66
Tabela 6 - Matriz Bidimensional, Anderson et al. (2011) ;Clark, Chopeta (2004); Clark, Mayer (2007)..	66

Lista de Abreviaturas

AUC – Area Under the Curve

BI – Business Intelligence

CRISP-DM - Cross Industry Standard Process for Data Mining

DM – Data mining

DMAIC - Define, Measure, Analyze, Improve, Control

DT – Árvores de Decisão

DW - Data Warehouse

EM –Escolha Múltipla

MLPE - Multilayer Perceptron

KDD - *Knowledge Discovery in Databases* (descoberta de conhecimento em base de dados)

NB - Redes Bayesianas

NN - Redes Neurais

OLAP - Online Analytical Processing

QA – Questões Abertas

ROC - Característica Operacional do Receptor

SEMMA - Sample, Explore, Modify, Model, Assessment

SVM - Máquinas de Vetores de Suporte

1. Introdução

1.1. Enquadramento e Problema

O processo de avaliação educacional é vasto e exige uma elevada responsabilidade, incluindo mecanismos próprios que são gerados dentro do próprio sistema de ensino, cujo objetivo fundamental é aferir a eficiência dos progressos de aprendizagem usados em cada situação concreta. Isto torna a avaliação num ponto importante de estudo, apostando na melhoria da qualidade da mesma e consequentemente do próprio ensino. Existem diversos formatos de questões de avaliação, como por exemplo Questões Abertas (QA), Escolha Múltipla (EM), entre outros. Na presente análise, as questões de múltipla escolha irão ser referenciadas por EM, sigla adotada por um guião utilizado para apoio, chamado *Guia de elaboração e revisão de questões e itens de Múltipla Escolha* (2011).

O problema central deste estudo passa por encontrar padrões e relações entre as variáveis consideradas com base na amostra de testes dos últimos três anos, realizados nas unidades curriculares de Excel Avançado e Introdução a Excel do ISCTE-IUL. Estes testes foram compostos por QA e de EM com quatro opções de resposta e serão o auxílio na procura de padrões e relações entre variáveis enquadradas em diferentes dimensões, aplicando DM.

Pretende-se ainda com este estudo chegar a resultados empíricos que indiquem quais os fatores mais influenciadores na classificação obtidos pelo discente nas questões de EM. Isto para que futuramente o docente na construção de exames, que contenha este bloco de questões, tenha em consideração determinados fatores de avaliação que possam vir a prejudicar e/ou beneficiar os alunos, caso não sejam elaborados com cautela.

Foi analisada uma amostra de testes de Excel Avançado, constituídos por dois blocos, uma de QA e outra de perguntas de EM do tipo resposta única. Estes testes serão analisados, e ser-lhe-ão aplicados várias experiências utilizando técnicas de DM, para que no fim se obtenham resultados, utilizados para análises, a fim de se descobrir quais dos fatores considerados estão na base da influência da classificação que o aluno obtém nos referidos exames. O conhecimento extraído irá servir de apoio para construção de futuros exames.

1.2. Objetivos

Pretende-se com esta dissertação definir variáveis que podem ser os principais fatores do problema, com base na análise de dados dos testes de Excel, nomeadamente das respostas a perguntas de EM. Com isso espera-se obter, utilizando as técnicas de modelação de DM, padrões relevantes entre as variáveis definidas que poderão tornar mais eficiente a construção de testes com este tipo de questões. Neste sentido, procura-se perceber, a partir das várias variáveis que serão consideradas como input, quais as que tem maior peso no resultado do bloco de EM, e até que ponto estas podem influenciar a nota obtida pelo discente, permitindo a construção de avaliações mais fiáveis e com menos enviesamento no resultado da avaliação.

Resumindo, este trabalho foca-se na realização de um estudo empírico em que se espera chegar a resultados explicativos das principais causas do insucesso nas questões de EM, ou seja, os fatores que influenciam as classificações obtidas neste bloco de questões do exame.

Isto para que seja possível cumprir um outro objetivo – o de auxiliar o docente com os resultados obtidos e com análise da mesma ter uma perceção de como poderá melhorar o seu desempenho na construção dos testes desta natureza, e consequentemente chegar a classificações satisfatórias.

1.3. Estrutura da dissertação

Esta dissertação inicia-se com a presente introdução e desenrola-se por mais dois capítulos, sendo estes ainda decompostos em subsecções.

No capítulo 2 é apresentado o resultado da pesquisa bibliográfica traduzida numa revisão da literatura, contextualizando o tema que se prende a este estudo e serão abordados aspetos sobre a avaliação académica de um modo geral. Seguidamente é analisada a Taxonomia de Bloom e a sua posterior revisão, ferramenta utilizada para classificação dos objetivos académicos consoante os diferentes níveis de conhecimento definidos. Numa outra subsecção do capítulo será explicado de forma pormenorizada as EM, que é o foco do estudo.

Na subsecção 2.4 são explicadas as principais utilidades do uso das folhas de cálculo, fazendo ponte com o meio académico, ilustrando alguns casos de exemplo do uso do Microsoft Excel nas avaliações e os resultados. Posteriormente serão apresentados casos de estudo de

exemplo, em que se aplicou análise de dados em testes com a mesma estrutura das que irão ser analisadas.

Após esta subsecção, é feito um levantamento do quadro teórico de referência na área de BI, bem como a aplicação de técnicas de DM e as suas diversas vertentes e aplicações. Finalmente, serão apresentados casos de aplicação de DM nas avaliações académicas.

Seguidamente na secção 3, é abordada a metodologia usada, assim como as ferramentas utilizadas e o detalhe de todo o desenrolar do trabalho realizado.

Por fim, a análise dos resultados obtidos após o estudo presentes na secção 4, seguidos das conclusões tiradas deste projeto.

2. Revisão da literatura

2.1. Avaliação académica

A avaliação pode ser caracterizada como uma forma de aferição da qualidade do avaliado. Numa avaliação integrada, a recolha de dados é feita a partir de diversas fontes, através de instrumentação diversa, e tem o contributo de professores, alunos e encarregados de educação. Neste sentido, Pacheco (2002) define avaliar como decidir na base de um juízo de valor, formulado sobre um processo de recolha de dados, em função de critérios bem definidos. Citando Fullan e Hargreaves (2001, p. 72), "o ensino não é a profissão mais antiga do mundo, mas é certamente, uma das mais solitárias". Deste modo, a formulação de critérios de avaliação exige a colaboração, o trabalho conjunto, a discussão, a problematização entre professores, e a assunção não artificial de competências pelos órgãos e elementos responsáveis na escola.

A complexidade da avaliação da aprendizagem exige um esforço conjunto de todos os atores educativos, principalmente quando se pressupõe que a escola existe para a promoção do sucesso educativo (Pacheco, 2002). Isto leva às seguintes questões: será possível realizar uma avaliação académica equitativa e que avalie de forma justa o conteúdo lecionado nas aulas? Pode-se ir além da memória do que se aprendeu? Requerendo, neste sentido, uma taxonomia que avalie os níveis os domínios de aprendizagem, alguns investigadores antigos propuseram várias (ex: Bloom et al., 1956; Gagné, 1968; Miller et al. 1978; Royer et al., 1993). Destas, a de Benjamin Bloom é das mais antigas, conhecidas e estudadas.

2.2. Taxonomia de Bloom

A Taxonomia de Bloom é uma estrutura de organização hierárquica de objetivos educacionais, resultado de investigação da equipa, que identificou três domínios de atividade de ensino ou aprendizagem: o cognitivo, com foco nas habilidades mentais (conhecimento); o afetivo, que se debruça sobre o crescimento em sentimentos ou áreas emocionais; e por fim, o domínio psicomotor que abrange as habilidades manuais ou físicas (Bloom et al., 1956). Apresentados os vários domínios da taxonomia, ir-se-á primeiramente abordar o domínio cognitivo, sendo pertinente para o tema em questão, visando o objetivo de avaliar a

aprendizagem e capacidade de análise sobre os conteúdos lecionados, de modo a auxiliar os docentes na construção de uma melhor avaliação pedagógica.

O domínio cognitivo da taxonomia original envolve seis categorias principais que são apresentadas numa hierarquia de complexidade e dependência, do mais simples ao mais complexo. Para ascender a uma nova categoria, é preciso ter obtido um desempenho adequado na anterior, pois cada uma utiliza capacidades adquiridas nos níveis anteriores, evidenciadas na Tabela 5. As categorias desse domínio são: Conhecimento; Compreensão; Aplicação; Análise; Síntese; e Avaliação.

Segundo Conklin (2005), a Taxonomia de Bloom e a sua classificação hierárquica dos objetivos de aprendizagem têm sido uma das maiores contribuições académicas para educadores que, conscientemente, procuram meios de estimular, nos seus discentes, o raciocínio e abstrações de alto nível (*higher order thinking*), sem distanciar-se dos objetivos educacionais previamente propostos, tendo auxiliado os educadores a criar eventos de aprendizagem significativos e resultados positivos nos alunos.

O contributo de Bloom e Krathwohl (1956) foi bastante notório, por trazer a possibilidade de padronização da linguagem no meio académico e, com isso, também novas discussões em redor dos assuntos relacionados à definição de objetivos educacionais. Neste contexto, instrumentos de aprendizagem puderam ser trabalhados de forma mais integrada e estruturada, inclusive considerando os avanços tecnológicos que podiam prover novas e diferentes ferramentas para facilitar o processo de ensino. Com o passar dos anos, novos conceitos, recursos e teorias foram incorporados no campo académico; avanços psicopedagógicos e tecnológicos ocorreram, o que levou à revisão dos pressupostos teóricos da Taxonomia original. Um grupo de especialistas (psicólogos, educadores, especialistas em currículos, testes, avaliação etc.) foram supervisionados por David Krathwohl, que participou no desenvolvimento da Taxonomia original no ano de 1956, e, no ano de 2001, o relatório da revisão foi publicado num livro intitulado “*A taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy for educational objectives*” (Anderson et al., 2005).

Na revisão destacaram alguns pontos considerados relevantes para os autores da nova versão da taxonomia, nomeadamente a divisão do conhecimento em dois tipos: (1) Conhecimento como processo e (2) conhecimento como conteúdo assimilado, como se pode ver na tabela 6, em anexo

da matriz bidimensional, a subdivisão da dimensão conhecimento. Também foi feita a reorganização da ordem categórica e a renomeação das categorias de substantivos para verbo, originando uma nova tabela da Taxonomia (Tabela 5, em anexo).

A matriz Cognitiva vs. Conhecimento, resultou da combinação dos processos cognitivos com os três níveis existentes da taxonomia original: fatorial, conceitual e processual. Não obstante a permanência destes níveis, foi ainda adicionado um novo nível – metacognição -, por Krathwohl e Anderson. Posteriormente Clark (2004), Clark e Mayer (2007), acrescentaram mais outros dois níveis de conhecimento: procedimentos e princípios, - tendo assim a matriz mais recente seis dimensões cognitivas.

Procedendo à descrição dos níveis da dimensão cognitiva apresentadas na Tabela 3 (anexo), tem-se a dimensão dos fatos - dados ou instância específica e única; dimensão conceitos - abrange uma classe de itens, palavras ou ideias que são conhecidos por um nome comum - inclui vários exemplos específicos e que compartilha características comuns. Existem dois tipos de conceitos: concreto e abstrato. Quanto ao domínio dos processos, caracteriza-se por ser um fluxo de eventos ou atividades que descrevem como as coisas funcionam em vez de como fazer as coisas. Normalmente existem dois tipos: os processos de negócios que descrevem fluxos de trabalho, e processos técnicos que descrevem como as coisas funcionam em equipamentos ou natureza. Relativamente aos procedimentos, definem-se numa série de ações passo-a-passo e decisões que resultam na realização de uma tarefa. Existem dois tipos de ações: lineares e ramificadas. A dimensão princípios, por sua vez, é um conjunto de diretrizes e regras, - inclui não apenas o que deve ser feito, mas também o que não deve ser feito. Permite fazer previsões e tirar implicações, e até inferir a causa de um fenómeno. Por fim, a dimensão metacognitivo que envolve o conhecimento cognitivo real, assim como a consciência da aprendizagem individual.

Com esta nova versão, foi atribuída maior flexibilidade ao conceito cumulativo e dependência entre as categorias, sabendo-se que os processos cognitivos diferem de disciplina para disciplina, possibilitando aos alunos aprenderem melhor num estágio mais elevado e depois serem capazes de entender os anteriores (Krathwohl, 2002). Salientando que com a abordagem da matriz criada na revisão provê um novo direcionamento para que educadores possam planejar melhor os seus objetivos de aprendizagem e direcionem, de forma coerente, clara e concisa, o seu processo de ensino.

Esta taxonomia irá contribuir para classificação das questões de EM de exames que virão a ser realizados, de modo a auxiliar o docente a classificar cada questão individualmente quanto ao nível cognitivo a que este requer do discente, assim como foi feito nas questões analisadas neste estudo. O enfoque nesta taxonomia, serve para evitar que o docente elabore questões que exijam conhecimentos que não vão de encontro ao que o aluno sabe, por isso a preocupação em abordá-la. Pretende-se que todo este processo de aprendizagem ocorra de forma justa, e equitativa, no sentido em que o aluno deverá estar devidamente preparado para as avaliações que lhe serão submetidas.

2.3. Questões de escolha múltipla

No seio académico existem inúmeros instrumentos que podem ser utilizados para avaliação dos alunos, sendo que as perguntas no formato de EM foram as consideradas para o presente estudo. O discente, durante o seu percurso académico, irá sempre deparar-se com perguntas neste formato, pois parte dos exames a que um aluno é submetido ao longo do seu percurso contém questões desta natureza.

As questões do formato de EM são consideradas questões objetivas, e podem ser de vários tipos, como de resposta única ou múltipla, de asserção ou negação (verdadeiro ou falso), entre outros, *Guia de elaboração e revisão de questões e itens de Múltipla Escolha* (2011). Porém, as questões que foram analisados neste trabalho são as de resposta única, em que a pergunta enuncia um problema e um leque de opções, do qual o aluno tem que escolher apenas uma opção, a que considera ser a correta.

Questões objetivas de EM são aquelas que já trazem enunciadas as possibilidades de resposta, de entre as quais o aluno escolhe a única que responde corretamente ao problema proposto. Uma questão objetiva não diz respeito apenas ao tipo de resposta que solicita. É objetiva pelo facto de o seu enunciado ser direto, limpo e o mais claro e preciso e também pela organização - pela forma como o problema é contextualizado e apresentado para a análise do aluno. A escolha do tipo de questão a ser formulada depende da natureza do conteúdo, da complexidade, da habilidade e do nível de competência que serão avaliados (Pinto, 2001).

A discussão entre a avaliação realizada através de questões de diferentes formatos, sendo as mais tradicionais os de QA e de EM, tem vindo a ser alvo de muita análise na esfera pedagógica,

tanto que se encontram na literatura vários estudos com opiniões divergentes face à adoção de um formato em detrimento do outro.

Em contraste, Becker e Johnston (1999), e Anderson et al. (2005) defendem que os dois formatos não medem uma mesma dimensão de conhecimento. Hancock (1994) por sua vez afirma que essa visão é parcialmente correta, pois para os níveis mais elevados da taxonomia da aprendizagem de Bloom, é preferível o uso de QA, dado que estas evocam com mais frequência o pensamento complexo, face às EM que frequentemente provocam o processamento cognitivo de baixo nível. Acrescentando ainda que, mesmo quando as pesquisas demonstram que os formatos de respostas variam em características cognitivas, psicométricas, custos de administração e pontuação, as implicações políticas permanecem ambíguas devido às frequentes discussões nessas dimensões. Partilhando da mesma opinião, Walstad e Becker (1994) afirmam que a partir de um ponto de vista económico sobre tempo e redução de custos, testes de EM devem ser preferidos.

É da opinião de alguns escritores que a construção de testes e seleção do formato deveria ser o alinhamento dos testes com os propósitos da educação. O teórico do currículo, Tyler (1994), por exemplo, afirmou que é importante "estabelecer um programa de testes que reflita fielmente os objetivos definidos pela escola". Similarmente, Traub (1993) propôs que a "escolha dos testes dependesse do que é medido, ou seja, a escolha deve depender da validade, não da dificuldade e da confiabilidade".

Contudo, estudos mais recentes continuam a explorar esta temática, tendo Rodriguez (2003) também descoberto a inconstância de resultado nestes formatos a partir de uma meta-análise, concluindo que quando os itens são construídos em ambos os formatos usando o mesmo enunciado, a correlação média entre eles é significativamente maior do que quando se utilizam itens com enunciados não equivalentes. A equivalência de construção, em parte, parece ser uma função do método de desenho do item ou da intenção do escritor do item. Por outro lado, Hohensinn e Kubinger (2011) realizaram um estudo em que aplicaram testes com dois formatos diferentes, QA e EM, sendo que os resultados obtidos indicam que os formatos de respostas têm um impacto distinto relativamente à dificuldade dos itens.

Com estas opiniões, todas de casos diferentes, pode-se aferir que não há uma estrutura padronizada que indique quando é que se deve aplicar o formato de QA ao invés de EM, e vice-versa, tendo neste caso o avaliador de saber o que pretende avaliar.

Acrescentando que as questões de EM são realmente ótimos instrumentos de avaliação, porém, se não forem elaborados com cuidado, os resultados podem não ser satisfatórios, reforçando o porquê de ter sido abordado a taxonomia na subsecção anterior. Isto é, pretende-se mostrar que tanto as QA como as de EM, formatos mais utilizados durante o ensino pedagógico, são bons instrumentos de avaliação. Porém, para a elaboração de ambas terá de se ter cautela, e garantir que o dado nas aulas e todo o processo de ensinamento, é o suficiente para responder às perguntas dos exames.

2.4. Avaliação de desempenho académico em cursos de MS Excel

O Microsoft Excel é um *software* que pertence à categoria designada por folhas de cálculo que permitem inserir, organizar e analisar informação, e ainda permite automatizar todo o trabalho de modo a efetuar cálculos e tarefas específicas. O utilizador pode guardar, manipular, calcular e analisar dados tais como números, texto e fórmulas (Caldeira, 2014).

Segundo Caldeira (2014), a criação de fórmulas para a execução de determinados cálculos pode ser uma tarefa complicada e demorada. Muitos dos tipos mais comuns e utilizados de fórmulas já estão pré-definidos no Excel e têm a designação geral de FUNÇÕES. As funções são fórmulas pré-definidas que fazem cálculos, utilizando valores específicos, através da chamada de argumentos, com sintaxe pré-definida (Caldeira, 2014). O Excel contém inúmeras funções, agrupadas em diversas categorias: estatística, matemática e trigonometria, financeira, base de dados, com datas e horas, lógica, procura de dados e texto.

Apresentado o leque de funcionalidades que as folhas de cálculo do Excel possuem, úteis para o mundo computacional, justificam o porquê de o Excel ser bastante usado no seio académico como instrumento de avaliação. Com os avanços tecnológicos, as instituições de ensino perceberam que usando de forma adequada estas inovações, em particular a ferramenta do Excel, traria uma mais-valia tanto para a instituição como para o aproveitamento dos alunos.

Acrescenta-se ainda que o ensino do Excel se massificou nos últimos anos (Treadwell et al., 2013), tendo as escolas a preocupação de investir no ensino das funcionalidades da mesma, segundo Sheu e Wong (2006, pag. 223) - “os alunos precisam possuir conhecimento interdisciplinar (folhas de cálculo) para competir em um mercado cada vez mais competitivo”. Por outro lado, Awasthi, et al., 2010, relatou que “as folhas de cálculo ainda são a ferramenta mais comum de orçamentação e previsão usada”. E foi neste sentido que Treadwell et al. (2013) sentiram a necessidade de avaliar as habilidades do Excel dos alunos, a fim de descobrir qual o nível de conhecimento que estes possuíam. Nos resultados denotaram-se que a média dos alunos que terminaram o exame foi de 33,48%, concluindo que as habilidades em relação às folhas de cálculo possuídas pelos alunos eram ainda de iniciantes, levando-os a apostar num curso intensivo de Excel, dada a extrema importância da ferramenta no quotidiano do mercado de trabalho.

Todos estes estudos revelam a potencialidade que a ferramenta Excel tem vindo a desencadear no seio académico.

2.5. Casos de estudo

Têm sido realizados vários estudos com o objetivo de comparar os resultados obtidos entre testes, que seguem a mesma estrutura dos da presente amostra, por forma a analisar as implicações da utilização de uma abordagem em detrimento da outra em diferentes casos, havendo desta forma um histórico de metodologias utilizadas para análise dos dados das avaliações nestes dois formatos, que serão abordadas nesta subsecção.

2.5.1. Análise de dados de avaliação académica

Na maioria dos estudos encontrados aplicaram-se abordagens estatísticas para análise das avaliações, tais como análise de fatores e modelos de equações estruturais, correlações, análise de modelos de variância e avaliação das estatísticas dos itens. Bennett et al. (1991) realizou estudos utilizando a análise fatorial confirmatória para testar o ajuste de um modelo de dois fatores onde cada formato de item marcava o seu próprio fator. Os resultados mostraram uma solução de um único fator para fornecer o ajuste mais simples em cada uma de duas amostras aleatórias, obtendo uma alta correlação (0.71 a 0.97) entre os dois fatores.

Buchweitz (1974), por sua vez, utilizou testes estatísticos sobre as amostras, utilizando a média das pontuações dos testes de QA e EM na disciplina de Física Geral, e verificou que a diferença de médias tende a aumentar a favor da EM ao nível de 1% à medida que níveis de comportamento mais complexos são exigidos. Embora a correlação entre as duas variáveis demonstra que não existia uma diferença significativa entre os resultados das QA e das EM. Isso para todos os itens e para todos os níveis de conhecimento, compreensão e avaliação separadamente, concluindo que as duas partes têm elementos em comum.

Outros autores usaram métodos de análise de variância para abordar a questão de equivalência de formatos. Helping (1981) realizou um estudo baseado nesta metodologia, concluindo que as pontuações foram mais elevadas no teste de EM. É de salientar que houve uma correlação significativa entre o formato do teste e a complexidade da questão (perguntas de conhecimento versus questões de conceito), isto é, as questões conceituais eram comparativamente mais fáceis face às de EM, sugerindo que o impacto dos itens QA e EM difere no que diz respeito ao desempenho dos avaliados.

Houve também alguns autores como Bridgeman (1992), e Hohensinn e Kubinger (2011), que utilizaram um método diferente - análise da teoria da resposta ao item (TRI) - para examinar o impacto dos formatos de resposta, concluindo que apesar das diferenças de formato substanciais observadas para itens individuais, as pontuações totais para os testes de múltipla escolha e de frase aberta demonstraram padrões de correlação notavelmente semelhantes.

Por fim, certos autores analisaram a equivalência de formatos, comparando as correlações de validade preditiva. Bridgeman et al. (1992) examinou os efeitos diferenciados do gênero interagindo com efeitos de formato. Estas questões são abordadas mais diretamente por Downing e Norcini (1998) que examinaram diferenças de predição devido ao formato e Ryan e Franz (1998) que avaliaram a pesquisa sobre as interações de gênero com o formato.

Resumindo os estudos feitos, pode-se concluir que na maioria dos casos foram utilizados métodos estatísticos para o estudo da relação existente entre os formatos de QA e de EM, e o que se denotou foi que, embora os testes de EM sejam as que a classificação é maioritariamente mais alta, não significa que estas são as mais fáceis e que exigem um nível de conhecimento menor que as de QA, pois na maioria dos estudos acima referidos concluiu-se que há, de fato, uma diferença entre os dois formatos, mas que esta não é significativa.

2.5.2. Metodologias usadas na construção dos testes dos casos revistos

Para os diferentes estudos revistos, foram utilizados procedimentos distintos para a construção dos testes e análise das mesmas. Lissitz e Hou (2008) realizaram um estudo, aplicando testes de Álgebra, Biologia, Inglês e Governo, contendo perguntas de QA e EM, em que os alunos tinham uma grelha para a inserção das respostas corretas nas de EM. Em ambas as unidades curriculares, os testes projetados tinham como base avaliar as mesmas expectativas nas áreas de conteúdo, portanto, o conhecimento e as habilidades de compreensão, tendo as perguntas quatro opções de respostas, tendo entre estas um ou dois distratores. Os distratores indicam as alternativas incorretas à resolução da situação-problema proposta. Essas possibilidades de resposta devem ser plausíveis, isto é, devem parecer corretas para os avaliados do teste que não desenvolveram a habilidade em questão (Haladyna, 2012).

Martinez (1991) desenvolveu no seu estudo testes destinados a vários domínios da Ciência, sendo estes contruídos de acordo com as especificações do NAEP (National Assessment of Educational Progress, 1988). Cada questão foi classificada por tópico (ex: ecologia), natureza da ciência (ex: conceção de uma experiência) e capacidade de pensar (ex: resolução de problemas), sendo cada um dos itens revistos por um painel externo de cientistas.

Os estudos revelam que existe uma vasta dimensão de variáveis de orientação para a construção das perguntas, como o conhecimento, compreensão, entre outros como se pode ver nos exemplos acima mencionados.

Como se pôde verificar, existem também inúmeros métodos para analisar os dados e estudar a relação existentes entre os dois formatos de testes, embora as mais usadas sejam as de estatísticas. No presente caso ir-se-ão utilizar as técnicas de DM, temática abordada no seção seguinte.

2.6. *Data mining* e extração de conhecimento

Nesta secção irá ser explicado como se desencadeia o processo de extração de conhecimentos com base em dados, que na pratica é o que foi feito neste trabalho de descoberta de conhecimento, de modo a que o utilizador perceba o mecanismo da mineração de dados.

Começando por abordar o conceito de BI, é um termo abrangente que agrega arquiteturas, ferramentas, bases de dados, aplicações e metodologias (Raisinghani, 2004). Um dos maiores benefícios de BI é o facto de disponibilizar acesso e manipulação de forma interativa (por vezes em tempo real) a dados, e facultar a gestores e analistas o acesso a análises mais controladas e detalhadas. O processo de BI, em traços gerais, baseia-se na transformação de dados em informação, posteriormente em decisões e por fim em ações (Turban et al., 2010).

Associada ao BI surge a necessidade de descoberta de conhecimento em Bases de Dados. Tal conceito enquadra-se perfeitamente no contexto de BI e é especialmente relevante tendo em conta o volume de dados de que as organizações dispõem, mas com o problema de, tipicamente, os dados se encontrarem dispersos, uma vez que respondem a diferentes tipos de necessidades e funções. Nesta designação inclui-se também todo o software para extração, transformação e povoamento de DW, OLAP, análise de dados, DM e visualização (Berson et al., 2001). Assim, uma das formas de organizar os dados para responder a solicitações no âmbito de BI é agrupá-los em DW, que consistem em repositórios de dados com diversas origens, organizados sob um esquema único e num único local de forma a facilitar a obtenção de conhecimento para apoio à gestão (Han e Kamber, 2011). Apesar destes tipos de ferramentas suportarem a tomada de decisão, são necessárias outras ferramentas adicionais para uma análise de dados mais aprofundada, entrando assim numa área mais específica, DM termo utilizado para descrever a descoberta de conhecimento a partir de grandes quantidades de dados (Turban et al., 2010).

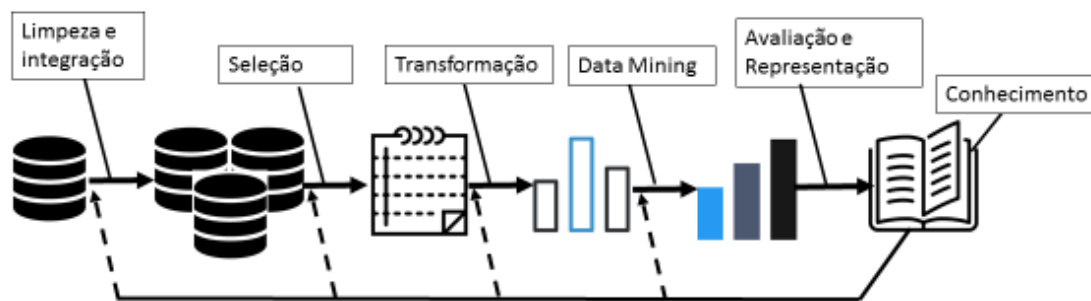


Figura 1 - Processo de KDD, adaptado de Han e Kamber (2006)

Este processo de KDD, numa primeira fase, passa pelo tratamento de dados, e seguidamente pela transformação dos padrões obtidos em conhecimento. É um processo com cinco fases que

utiliza métodos de DM (numa das cinco fases do processo) para extrair conhecimento através da identificação de padrões, de acordo com a especificação e os limites da fonte de dados. Essas fases correspondem a: (1) Limpeza e integração, (2) seleção, (3) transformação, (4) DM, (5) avaliação e representação, e por último (6) o conhecimento que é extraído de todo este processo - como se pode observar na Figura 1.

Contudo, dado o uso atual mais lato do termo DM e, tal como utilizado por Han e Kamber (2011), o termo DM irá futuramente ser usado como um sinónimo do processo da descoberta de conhecimento. Este não é mais do que um processo que utiliza um conjunto de técnicas estatísticas, matemáticas, inteligência artificial e aprendizagem máquina (*machine learning*) de forma a extrair e identificar informação útil, normalmente, num DW transformando-a posteriormente em conhecimento (Brito et al., 2006). Por outro lado, Turban et al., (2010) definem DM como sendo um processo de identificação matemática de padrões, normalmente através de grandes quantidades de dados.

O Grupo Gartner define DM como sendo um processo de descoberta de correlações significativas, padrões e tendências através de grandes quantidades de dados, armazenados em repositórios, utilizando tecnologias de reconhecimento de padrões, bem como técnicas estatísticas e matemáticas.

Han et al. (2011) preferem manter uma visão mais ampla, e definem DM como sendo um processo de descoberta de padrões e conhecimento relevantes com base em grandes quantidades de dados. As fontes destes dados incluem bases de dados, DW, Internet e outros repositórios de informação ou dados que são transmitidos para o sistema de forma dinâmica. Com a massificação do armazenamento de dados rapidamente os tempos que se seguiram ficaram marcados como a década da informação. Este autor ressalta que, atualmente e com a explosão das unidades de armazenamento cada vez maiores e mais eficientes, há condições para acreditar que nos encontramos na era dos dados; dados esses que carecem de ferramentas poderosas e versáteis para que possam extrair o máximo de informação possível e a transformem em conhecimento estruturado.

2.6.1. Métodos de abordagem em *Data mining*

Segundo Azevedo e Santos (2005), quando o processo de DM é enquadrado no contexto de uma metodologia, este torna-se mais fácil de compreender, implementar e desenvolver. Atualmente encontram-se disseminadas e bem definidas várias metodologias para o desenvolvimento do DM, como o CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify Model, Assessment) e DMAIC (Define Measure Analyze Improve Control), e o KDD (Knowledge Discovery in Databases). O SEMMA e o CRISP-DM (Figura 2). são as metodologias em que mais se têm desenvolvido projetos nesta área (Nogueira, 2014).

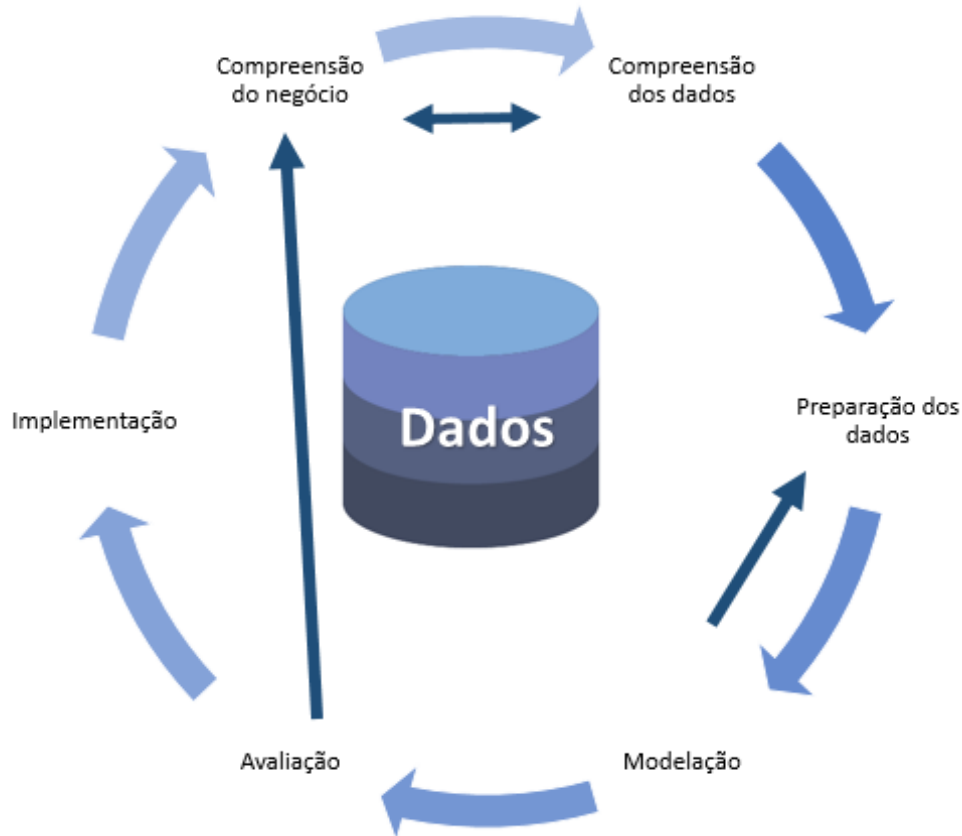


Figura 2 - Fases da metodologia CRISP-DM, adaptado de Chapman et al. (2000)

A metodologia **CRISP-DM** foi concebida em finais de 1996 e o seu desenvolvimento foi motivado pelo interesse crescente e generalizado, por um lado pelo mercado de DM, e por outro,

pelo consenso de que a indústria necessitava de um processo padronizado (Wirth, 2000). Os fundamentos desta metodologia, para além dos princípios académicos e teóricos, baseiam-se na prática e na experiência daqueles que desenvolvem de facto projetos de DM. O conhecimento prático foi assim incorporado de forma a dar resposta aos requisitos dos utilizadores, não se centrando unicamente na tecnologia, mas antes na resolução de problemas do negócio (Han et al., 2011). A metodologia CRISP-DM é descrita em termos de um processo hierárquico, com um ciclo de vida que se desenvolve em seis fases:

1. Compreensão do negócio: fase para compreender os objetivos do projeto e os requisitos de uma perspetiva de negócio, com o objetivo de converter esse conhecimento numa definição do problema de DM e num plano preliminar desenhado para atingir os objetivos.
2. Compreensão dos dados: fase de análise e compreensão dos dados, onde se identificam problemas de qualidade de dados e/ou se detetam subconjuntos interessantes de dados para formar hipóteses sobre a informação oculta nos dados.
3. Preparação dos dados: fase de construção do conjunto de dados finais para alimentação das ferramentas de modelação. As tarefas incluem selecionar os atributos, transformar e limpar os dados para utilizar nas ferramentas de modelação.
4. Modelação: fase de seleção e aplicação das técnicas de modelação, onde os parâmetros são calibrados para valores ótimos.
5. Avaliação: fase onde o principal objetivo é determinar se há algum problema de negócio importante que não foi suficientemente considerado. No final desta fase, a decisão sobre a utilização dos resultados de DM deve ser alcançada.
6. Implementação: esta fase pode ser simples, tal como a geração de um relatório, ou tão complexo como a implementação de um processo de DM repetível para avaliar os modelos e rever os passos executados para criá-lo, de modo a garantir que o modelo alcance adequadamente os objetivos propostos.

As fases não têm uma sequência fixa, dependendo do resultado e do desempenho das outras fases ou das tarefas particulares de determinada fase (Chapman et al., 2000).

Por outro lado, existe a metodologia **SEMMA** Sample, Explore, Modify, Model, Assessment, que foi desenvolvida pela empresa SAS (http://www.sas.com/pt_pt/home.html), cuja área de negócio é a Estatística, Análise de Dados, *Business Intelligence*, DM e Suporte à Decisão (Turban et al., 2010.)

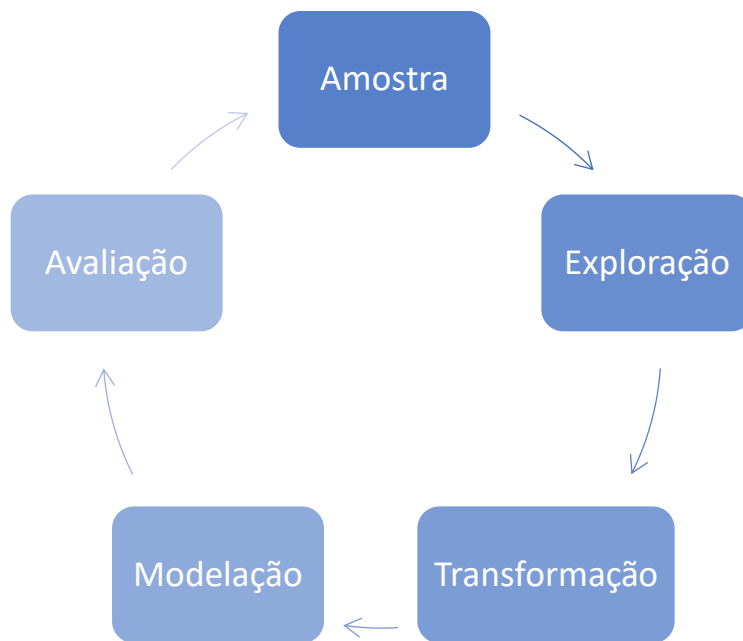


Figura 3 - Processo de Data Mining: SEMMA, adaptado de Turban et al. (2011)

Resumidamente, é possível apresentar esta metodologia como um processo com 5 fases que se inicia com uma amostra (*Sample*) representativa dos dados à qual se aplicam técnicas estatísticas de exploração e de visualização dos dados (*Explore*). Posteriormente são seleccionadas e transformadas as variáveis (*Modify*) consideradas mais significativas (as variáveis que sobressaíram na fase anterior), as que são mais relevantes em termos de projeto, e sobre as quais se constroem o modelo (*Model*) (aplicam-se algoritmos no sentido de alcançar os objetivos) e por fim avalia-se o modelo (*Assess*). Cada uma das etapas é distinta e corresponde a um ciclo, e as suas tarefas internas podem ser executadas repetidamente sempre que necessário, i.e., pode-se atualizar e ajustar quando surgir nova informação.

Só se abordou as metodologias SEMMA e CRISP-DM pelo facto de serem as mais usuais em projetos de DM, como foi mencionado anteriormente, embora pode-se depreender que em termos de processos para desenvolvimento de um projeto de DM a metodologia CRISP-DM é mais

completa que a SEMMA, pela incorporação das fases de Compreensão do Negócio e Implementação. Ressaltando que neste trabalho foi adotada a CRISP-DM por ser considerada mais completa quando comparada com o SEMMA, e pela notoriedade que tem vindo a ganhar nesta área, sendo inclusive considerada como popular no desenvolvimento de projetos desta natureza (Mendes, 2011).

2.6.2. Técnicas de Modelação em *Data mining*

Existem inúmeras técnicas para a modelação em DM usadas para diferentes propósitos e objetivos. Através da Figura 4 é possível compreender a variedade de alguns dos seus métodos e a sua relação. Um dos objetivos, a verificação, testa a hipótese do utilizador enquanto a descoberta procura novos padrões. O objetivo de descoberta por sua vez subdivide-se em dois - previsão e descrição. A previsão assume uma procura constante de padrões que permitam prever situações/comportamentos futuros e a descrição assenta numa procura de padrões que representem o conhecimento de forma compreensível.

Para o objetivo de descrição há vários métodos, nomeadamente: segmentação, sumarização, deteção de desvios, visualização e síntese linguística, entre outros. Por outro lado, no objetivo de previsão, a problemática assenta em dois pontos essenciais: a regressão onde se pretende encontrar uma função desconhecida cuja variável dependente tem um domínio de valores reais e a classificação onde se pretende encontrar uma função que faça o mapeamento dos dados em classes pré-definidas (Maimon & Rokach, 2010; Fayyad et al., 1996).

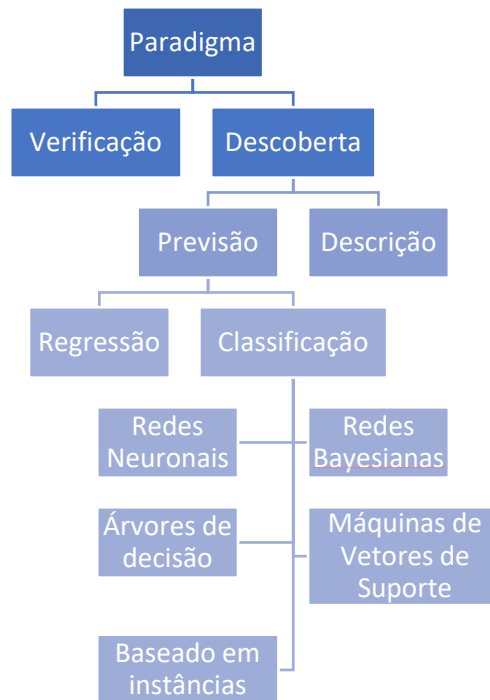


Figura 4 - Taxonomia do *Data Mining*, adaptado de Maimon & Rokach (2010)

Os algoritmos de DM podem dividir-se em quatro (4) grandes categorias: classificação, segmentação, associação e descoberta de seqüências. No entanto, poderão existir outras ferramentas de análise de dados, como a visualização, regressão e análise de séries temporais (Turban et al., 2010), podendo-se descrever sucintamente a taxonomia do seguinte modo:

Classificação - Trata-se do tipo de problema mais frequente em DM (Drozdek, 2005) e consiste na análise das características de um objeto e, através de um classificador¹, à associação dessas características a classes predeterminadas. O conceito assenta na esperança de que o modelo possa ser utilizado para prever classes de registos não classificados. Algumas das ferramentas mais utilizadas são as redes neuronais, árvores de decisão e outras regras sem uma estrutura definida. As redes neuronais são inspiradas na fisiologia do cérebro e nas complexas redes neuronais biológicas.

¹ Um **Classificador** classifica um conjunto de dados num número de possíveis categorias, definidas à priori pelo analista

À semelhança das redes biológicas, a técnica de redes neuronais é composta por um conjunto de unidades simples de processamento designadas “nós”, que interligados formam uma rede de nós e que em conjunto aumentam a capacidade computacional da unidade (Liao et al., 2012). Por sua vez, as árvores de decisão classificam os dados num número finito de classes com base em valores das variáveis de entrada.

Segmentação (clustering) - consiste na identificação de um conjunto finito de categorias ou *clusters* usados para descrição dos dados de acordo com uma métrica, segundo a qual se procura maximizar as semelhanças e minimizar as diferenças entre si. A aprendizagem neste modelo é do tipo não supervisionado, o que corresponde, na prática, a inexistência de qual a informação inicial no sistema - assume-se que este aprende por si mesmo. Nesta técnica, a pesquisa é realizada sobre os dados tendo em conta todas as características de cada dado.

Descoberta de sequências - Este consiste na identificação de associações ao longo do tempo. Quando uma determinada informação é facultada, pode realizar-se uma análise temporal, no sentido de verificar determinados comportamentos ao longo do tempo. Este processo fornece um conjunto de informação considerável e que pode ser utilizado, a título de exemplo, para o aumento de vendas ou deteção de fraudes.

Associação - estabelece relacionamentos entre registos que se encontram no mesmo registo.

Visualização - possibilita a representação gráfica dos resultados (finais ou intermédios) de DM recorrendo a formas visuais facilmente perceptíveis. O objetivo da visualização consiste em descrever as informações complexas através de diagramas, permitindo uma melhor representação de padrões e tendências. Quanto melhor for a descrição de um conjunto de dados, maior é a possibilidade de compreender o domínio em que está inserido.

Regressão – É uma técnica utilizada para modelar uma variável numérica com base num conjunto de dados de entrada caracterizados por várias variáveis, tendo como fim a previsão.

E como se pode observar pela Figura 4, existem inúmeros modelos de DM destinados a resolução de classificação, sendo este o problema do presente caso de estudo, nomeadamente as árvores de decisão (DTs), as redes bayesianas (NB), as máquinas de vetores de suporte (SVM) e as redes neuronais (NN).

As árvores de decisão (DTs) podem ser definidas como uma estrutura de dados recursivamente com nós folha - que indicam uma classe -, ou nós de decisão que contém um teste sobre o valor de um atributo (Dos Santos Rodrigues, 2005). A cada um desses atributos existe uma ligação a um novo ramo, desencadeando uma outra árvore de decisão designada de subárvore, tendo esta a mesma estrutura de uma árvore. Segundo Dos Santos Rodrigues, as árvores de decisão dividem o espaço de descrição do problema em regiões disjuntivas, isto é, um exemplo é apenas classificado por apenas um único ramo da árvore. É um método de classificação supervisionada, onde uma variável dependente é explicada à custa de n variáveis independentes, medidas em qualquer escala.

O algoritmo das redes bayesianas (NB) é um simples classificador probabilístico que calcula um conjunto de probabilidades contando a frequência e as combinações de valores num dado conjunto de dados. O algoritmo usa o teorema de Bayes e assume que todos os atributos são independentes do valor da variável de classe. Esta suposição de independência condicional raramente é verdadeira em aplicações do mundo real, daí a caracterização como Naive (simplório), ainda que o algoritmo tenda a executar bem e aprenda rapidamente em vários problemas de classificação supervisionados, (Dimitoglou, Adams e Jim, 2012).

Conforme afirma Cortes e Vapnik (1995), as máquinas de vetores de suporte são um método de aprendizagem supervisionado utilizado para a classificação e análise de regressão que encontra o hiper plano ótimo (função de decisão linear), que tem a margem máxima entre os vetores de duas classes separáveis (chamados vetores de suporte). O SVM encontra assim o melhor hiperplano de separação linear, relacionado a um conjunto de pontos vetoriais de suporte, no espaço de recurso.

Por último, existem as redes neuronais que são modelos simplificados de cérebro humano - é um processador paralelo composto por simples unidades de processamento (neurónios ou nodos), no qual o conhecimento é armazenado nas conexões (ligações) entre os nodos. Os nodos encontram-se interligados numa estrutura de redes. Existem vários tipos de arquitetura, no entanto, neste trabalho, usou-se o *Multilayer perceptron* – a topologia mais popular e usada da rede neuronal.

2.7. *Data mining* na avaliação acadêmica

Os campos de BI/DM que visam extrair conhecimento a partir de dados brutos, como já foi referido anteriormente, oferecem ferramentas úteis automatizadas que podem e têm auxiliado o domínio acadêmico, no sentido em que fornecem informações e resultados úteis para a tomada de decisão neste ramo.

Cortez e Silva (2008) desenvolveram um caso que serve de exemplo da utilização de DM no meio acadêmico, abordando a predição da classificação dos alunos de duas turmas básicas (Matemática e Português) utilizando dados demográficos, sociais e outras relacionadas como base do estudo. Para o tratamento de dados, foi utilizada a classificação da regressão binária de cinco níveis, e ainda quatro métodos de DM, nomeadamente: as árvores de decisão (DT), as redes neuronais (NN), máquinas suportadas por vetores (SVM) e as florestas aleatórias (RF). Foram também exploradas seleções de entradas distintas (por exemplo, com ou sem classificações anteriores); os resultados obtidos revelam que é possível obter uma alta precisão preditiva, desde que se conheçam as classificações do primeiro e/ou segundo ano escolar. No entanto, uma análise do conhecimento proporcionada pelos melhores modelos de previsão demonstrou que, em alguns casos, existem outras características relevantes, tais como: escolaridade (por exemplo, número de ausências, motivo para escolher a escola, por exemplo, idade dos pais, trabalho dos pais e educação) e sociais (por exemplo, sair com amigos, consumo de álcool).

Por outro lado, Coutinho e Padilha (2016) também desenvolveram um caso de estudo neste âmbito, cujo propósito é de analisar o desempenho de alunos numa avaliação contínua (30% para trabalhos e 70 % para testes) por meio de redes Bayesianas, sendo esta um dos métodos de DA (ver figura 2). Diante dos resultados concluiu-se que mesmo que os trabalhos (30%) tenham pesos inferiores em comparação com as provas (70%), estas influenciam fortemente para a avaliação dos alunos. Além disso, foi possível identificar, variáveis que influenciam e sinalizam o futuro do desempenho de alunos na disciplina de Lógica Aplicada à Computação no que diz respeito a aprovações, reprovações e desistências. Baradwaj e Pal (2012) tentaram prever o sucesso escolar dos alunos utilizando árvores de decisão. Para facilitar o uso e interpretação do algoritmo de classificação, todos os atributos são categorizados antes da análise. O modelo

criado deteta os alunos com grande probabilidade de desistir no curso, o que permite a intervenção dos professores antes do acontecimento.

Vandamme et al. (2007) procuram classificar antecipadamente os alunos que entram na instituição com o risco de desistirem de curso. São comparadas três técnicas diferentes para classificar os alunos em três classes diferentes (baixo risco de desistir, médio risco e alto risco). As técnicas usadas são redes neuronais, árvores de decisão e uma análise discriminante linear. Os dados utilizados são divididos em 30% para validação e 70% para treino. A validação dos resultados, para todas as técnicas, fica entre os 40% e os 50%, ou seja, o algoritmo treinado consegue classificar corretamente entre 40 a 50% dos estudantes, uma percentagem que não é muito satisfatória.

Similarmente, Amorim et al. (2008) desenvolveram um projeto cuja finalidade é de construção de um sistema de previsão de evasão escolar, fazendo com que a Instituição de Ensino Superior possa tomar as devidas providências para tentar reverter a situação, comparando o desempenho de três classificadores, nomeadamente a NB, SVM e DT, a fim de atestar quais podem apresentar os melhores resultados para o problema em questão. Concluiu-se com o estudo que com as técnicas de *machine learning* têm vindo a ser ferramentas úteis para a perceção e tomada de decisões estratégicas no tratamento de problemas relacionados com a evasão escolar.

Apresentados os casos de exemplo, pôde-se aferir que as técnicas de DM no âmbito académico têm vindo a ser bastante úteis para a tomada de decisão, no sentido em que possibilitam elaborar estudos com um leque de ferramentas e métodos, permitindo assim chegar às causas de possíveis problemas, de modo a melhorar futuramente o processo de ensino. No entanto, das pesquisas efetuadas, não foi encontrado nenhum caso parecido com o que se pretende estudar nesta dissertação. Este estudo passa por aplicar as metodologias de DM focadas sobre as respostas a perguntas de EM de testes de Excel, com vista a descobrir padrões que possam estar a influenciar melhores resultados neste teste face às de QA.

3. Metodologia de investigação

Pretende-se, com este capítulo, abordar todo o trabalho realizado com este projeto, iniciando com uma breve contextualização do problema seguido da descrição do caso em estudo, revelando o método aplicado na abordagem ao tema da dissertação, assim como o universo em que vai incidir a análise. Serão indicadas as técnicas utilizadas, bem como o modo como irá ser feita a recolha de dados e o posterior tratamento, explicitar as ferramentas utilizadas para análise da amostra, e por fim as respetivas interpretações dos resultados obtidos.

No seio académico, a avaliação é das componentes que exige grande esforço do corpo docente, tanto para a estipulação de um planeamento que vai de encontro ao que é lecionado, como para a elaboração de instrumentos de avaliação justos que permitam avaliar no discente o aprendido nas aulas.

Entre esses instrumentos existem os testes constituídos por questões no formato EM e QA, que serão os analisados no caso em estudo. Sabe-se que existem vários estudos realizados que apontam as vantagens de um formato face ao outro. Neste caso, o que se pretende é das várias variáveis input consideradas analisar quais as que influenciam os resultados obtidos nos de EM, tendo por base modelos de classificação.

Existem vários modelos de classificação, apontados anteriormente no estado da arte: as árvores de decisão (DTs), as redes bayesianas, as redes neuronais (NNs), e as máquinas de vetores de suporte (SVMs). O último modelo constitui uma técnica de aprendizado que vem recebendo crescente atenção da comunidade de Aprendizado de Máquina, os resultados da aplicação dessa técnica são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado como as de Redes Neuronais (NNs) (Lorena e Carvalho, 2007).

Esses modelos encontram-se implementados na biblioteca *rminer*, sendo esta instalada em ambiente R, *open source* e multiplataforma (Windows, Linux e Mac OS). O ambiente R será utilizado para a realização da componente experimental.

3.1. Trabalho realizado

Tendo por base o estado da arte apresentado, a metodologia escolhida para a sistematização e descoberta de conhecimento foi o CRISP-DM. A escolha recaiu sobre esta metodologia por se tratar de um modelo padrão e não proprietário, e por possuir todas as fases que possibilitam a descoberta do conhecimento por detrás do problema em estudo.

3.1.1. CRIPS-DM – compreensão do negócio

A metodologia CRISP-DM propõe seis fases, iniciadas com a fase de **compreensão do negócio** (*business understanding*). Numa fase preliminar do projeto, coube à autora do presente estudo perceber o que se pretendia com este trabalho de forma detalhada, para que fosse possível prosseguir com o mesmo.

Considerando-se que esta fase é crucial para direcionar a investigação para o rumo acertado, seguiu-se as fases do CRISP-DM, de modo a organizar as etapas por que se delineiam este vasto processo de descoberta de conhecimento.

Posto isto, teve que se perceber como os testes se encontravam estruturados, a que anos diziam respeito as amostras que se iriam recolher e o que se pretendia com a análise das mesmas. Ou seja, teve que se perceber a amostra que foi disponibilizada, o que era importante recolher da mesma, e quais os objetivos espectáveis após a realização do estudo.

Portanto, no presente trabalho foram analisados dados relativos a exames da unidade curricular Excel, Introdução e Avançada, realizados no ISCTE, referentes aos anos 2012 a 2016. A estrutura destes exames era dividida em dois blocos, uma de QA e a outra de questões de EM.

Estes testes foram realizados à mão (papel e caneta), em que os avaliados teriam que utilizar fórmulas objetivas do Excel para responder às questões na parte aberta, sendo diferente do que normalmente é pedido nestes tipos de respostas. Ou seja, não há respostas de texto livre, com descrições e justificações.

No tipo de questões de EM analisadas, o avaliado tem de escolher de entre o conjunto de opções a que considerar correta, isto é, questões de EM do tipo resposta única. O bloco de questões por vezes pode ter por base imagens de tabelas ou de células da folha de cálculo preenchidas, em que o docente pretendia que o aluno seleccionasse qual o resultado esperado para cada pergunta em particular.

O número de perguntas de cada bloco era quase o mesmo na maioria dos testes, porém havia alguns casos em que as de QA eram mais do que as de EM. No entanto, em grande parte da amostra existiam testes com 6 questões de EM com quatro opções, com uma cotação de 1 valor por cada questão, sendo descontados 0,25 valores no caso de o aluno escolher uma das opções erradas, e o bloco de QA com 7 questões com cotação de 2 valores cada.

Após a recolha de toda informação, teve-se que realizar a tarefa de compreender cada variável *input* individualmente, perceber a particularidade de cada uma, e até que ponto seria útil ou não as consideradas para o estudo, tendo que conciliar as mesmas com o objetivo do próprio problema, que é perceber quais os principais fatores que influenciam a classificação obtida pelo discente nos testes de EM.

3.1.2. CRIPS-DM - compreensão dos dados

Na fase **compreensão dos dados** (*data understanding*) é feita a aquisição dos dados iniciais, e são descritos, explorados e verificados, avaliando a qualidade dos mesmos. Inicialmente fez-se a recolha e interpretação dos dados utilizando a ferramenta *Microsoft Access*. Optou-se por se usar esta ferramenta para uma recolha de dados de forma organizada, e posteriormente exportar os mesmos agrupando toda a informação, recolhida manualmente dos testes disponibilizados por parte do ISCTE, para o *Excel*.

Tendo os dados no *Excel*, foram criadas algumas outras variáveis a partir das que já existiam, tais como a nota final obtida no exame e os resultados discriminados, de forma a que se facilitasse o estudo das mesmas utilizando fórmulas disponíveis na ferramenta *Excel*, enriquecendo o conjunto de dados já existentes.

Como neste tópico se pretende perceber bem os dados, fez-se uma análise estatística simples para se compreender como os dados de algumas variáveis consideradas mais relevantes encontram distribuídos. As áreas dos cursos encontram-se divididos em três: “Tecnologia”, “Gestão” e “Sociologia” - sendo que a primeira ocupa quase toda a amostra dos testes com 90% de exames desta área, a segunda com 6% da amostra, e os restantes 4% relativos à área de Sociologia.

Quanto às unidades curriculares, tem-se apenas duas na amostra: “Excel Avançado” e “Introdução a Excel” - a primeira ocupando 95% da amostra e a segunda

com apenas 5%. Relativamente aos “Blocos de Matéria”, tem-se as “Fórmulas”, que ocupam a maior parte da amostra com 35%, as “Funções” com 25%, as “Funções Lógicas” e os “Sobre a aplicação” com 17%, as Tabelas e as Funções de Texto com 3% e 2% respetivamente. Desta forma consegue-se perceber quais os blocos mais frequentes nos exames face aos outros.

Quanto ao grau de dificuldade, a maioria das questões da amostra são de nível “Fácil”, ocupando 49% do total, o nível “Médio” com 39% de frequência, sendo os restantes das perguntas referentes ao nível “Elevado”. Da amostra ainda é relevante salientar que quase metade das questões contém apenas uma opção próxima, 40% com nenhuma, e os restantes 10% tem dois e três opções próximas.

Acrescentando ainda que dos 16 diferentes testes que constituem a amostra, o teste 12 ocupa 35% da amostra, face aos restantes que ocupam no máximo 9% individualmente. Este facto demonstra que o teste 12 foi realizado por vários alunos, e que foi aplicada em mais do que uma fase de avaliação durante o período letivo, o que se pôde perceber ao analisar cuidadosamente a amostra. Sendo esta a fase de descrição das variáveis, ir-se-á começar por analisar as variáveis por grupos. Sendo algumas variáveis referentes ao aluno, outras ao teste e outras à unidade curricular, a descrição será feita separadamente de maneira agrupada.

3.1.2.1. Dados caracterizadores do aluno

Começando pelo NumAluno, que diz respeito ao número de aluno, é a identificação que diferencia os alunos uns dos outros. Ainda relativo ao aluno tem-se o AlunoTeste que diz respeito ao número do teste que o aluno realizou. Isto para que seja possível uma análise mais concreta, sabendo-se qual foi o teste que o aluno fez, sabe-se quais os tipos de perguntas que o constituíam, e a estrutura do mesmo. E por último a variável Curso, que indica o curso frequentado pelo discente.

3.1.2.2. Dados caracterizadores da unidade curricular

Quanto à unidade curricular, existe apenas uma variável que a caracteriza, a NomeUC. Esta variável dá a informação relativa ao nome da unidade curricular frequentada pelo aluno, visto que existem duas possibilidades: “Excel Avançado” e “Introdução a Excel”.

3.1.2.3. Dados caracterizadores do teste

Neste grupo o número de variáveis é mais vasto, pois é onde está quase toda informação relevante para o estudo. Começando pela variável NumTeste, esta indica qual o teste realizado pelo aluno. Todos os testes da amostra foram identificados com um número, consoante a ordem de inserção dos mesmos na base de dados. Esse número serve para identificar o teste e quais as perguntas referentes a cada um.

As variáveis Mês e Ano dizem respeito ao próprio teste realizado pelo aluno e designam o Mês e o Ano em que o teste foi feito, salientando que a variável Ano é relativo aos últimos três anos.

Relativamente ao BlocoMateria e GrauDificuldade, ambas dizem respeito aos testes, sendo que o Bloco representa os vários tópicos de matéria que as perguntas podem abordar. Estes podem ser categorizados em “Fórmulas”, “Funções básicas”, “Funções de texto”, “Funções Lógicas”, “Sobre a aplicação” e “Tabelas”. Quanto ao grau de dificuldade da pergunta, pode assumir três valores: “Fácil”, “Médio” e “Elevado”. É de notar que ambas as variáveis foram usadas para classificar apenas as questões de EM.

No seguimento da análise deparou-se com obstáculos no tratamento das variáveis TextoAlínea e TipoAlínea. A primeira diz respeito ao texto das opções de resposta, enquanto que o segundo é referente ao tipo das opções, que se encontra categorizada em Certa, Próxima e Distrator.

Para tratar o texto teria de se usar *Text Mining*, uma técnica diferente de *Data Mining*, e como não era o que se pretendia com este projeto, optou-se pela exclusão da variável das experiências que seriam realizadas. Em relação aos tipos, optou-se por esmiuçar a informação disponível e criar novas variáveis a partir destas. Decidiu-se criar duas novas variáveis, NumProxima e NumDistrator, que para cada pergunta indicasse quantas opções próximas e distratores esta possuía, pois certa só pode existir uma, visto que se tratam de perguntas de escolha única.

As variáveis resultado, ResCerta e ResSelecionada, representam o resultado obtida numa pergunta individualmente: a alínea certa para a determinada questão e qual foi a resposta selecionada pelo aluno na questão, respetivamente. Enquanto que a CotPergunta diz respeito à cotação que o aluno terá caso acerte na resposta à questão.

Quanto à NotaEM e PesoEM, a nota diz respeito à classificação obtida pelo discente neste bloco do exame, enquanto que o peso é a cotação que o aluno obterá se acertasse todo o bloco, ou seja, é o peso desse bloco no exame, que pode ser 4 (contém 4 perguntas), 5 (contém 5 perguntas), 6 (contém 6 perguntas) ou 8, (contém 8 perguntas) com uma cotação de 1 valor para cada uma das perguntas.

Relativamente às variáveis NotaQA e NotaUC, também referentes ao teste, ambas indicam as notas obtidas pelos discentes no bloco de QA e na unidade curricular em si, respetivamente. Ainda referente aos testes tem-se a variável EnunTipo, que diz respeito ao tipo do enunciado de uma determinada questão, podendo ser de aplicação, mais teoria ou de resolução de problemas, ou seja, mais prática. Por fim tem-se o Enunciado, que possui o mesmo problema dos textos das opções, ou seja, teriam de ser estudadas usando a técnica de *Text Mining*; neste caso optou-se por usar as próprias funções do Excel, e criar duas variáveis a partir daquela, o NumPalavras e NumCarateres, que indicam o número de palavras e de caracteres que o enunciado da questão contém.

Nesta secção foi feita a descrição de todas as variáveis - as que já existiam e as que foram criadas, separando-as em grupos diferentes para que fosse possível uma melhor compreensão.

Seguidamente ir-se-á proceder à preparação das mesmas.

3.1.3. CRISP-DM - preparação dos dados

Nesta fase será feita a **preparação dos dados** (*data preparation*). Estes serão selecionados e preparados para a fase de modelação. Entre as tarefas destacam-se a limpeza, formatação e integração dos dados. O R, ferramenta que irá ser usado para construir os modelos de DM, necessita de um *input* único - um ficheiro .csv neste caso, em que cada linha corresponde a um registo independente de uma ocorrência. Isto é, cada linha diz respeito a uma resposta de uma determinada pergunta de um determinado aluno a um determinado exame.

Após ter todas as variáveis, procedeu-se à classificação das mesmas (Tabela 1), dividindo-as em diferentes categorias: numéricas (*numeric*), inteiros (*integer*), caracteres (*character*) ou categóricas (*factor*), à semelhança do que foi feito num estudo da mesma natureza realizado por Moro et al. (2017).

Tabela 1 - Análise das variáveis

Nome das variáveis	Tipo	Estado
NumAluno	factor	Excluído
AlunoTeste	factor	Incluído
Curso	factor	Excluído
AreaCurso	factor	Incluído
NomeUC	factor	Incluído
NumTeste	factor	Excluído
Mes	integer	Incluído
Ano	factor	Incluído
numTestesAno	numeric	Incluído
NumPergunta	integer	Excluído
BlocoMatéria	factor	Incluído
GrauDificuldade	factor	Incluído
TextoAlínea1	character	Excluído
TipoAlínea1	factor	Transformado
TextoAlínea2	character	Excluído
TipoAlínea2	factor	Transformado
TextoAlínea3	character	Excluído
TipoAlínea3	factor	Transformado
TextoAlínea4	character	Excluído
TipoAlínea4	factor	Transformado
NumDistrator	factor	Excluído
NumProxima	factor	Incluído
Imagem	factor	Incluído
resultado	factor	Target
ResCerta	factor	Excluído
ResSelecionada	factor	Excluído
CotPergunta	numeric	Excluído
NotaEM	numeric	Incluído
PesoEM	factor	Incluído
NotaQA	numeric	Incluído
NotaUC	numeric	Incluído
Enunciado	character	Excluído
EnunTipo	factor	Incluído
NumPalavras	numeric	Incluído
NumCarateres	numeric	Incluído

Sendo esta a fase que antecede a modelação dos dados, a tarefa será escolher todas as variáveis a serem consideradas nas experiências e, como se pode observar pela coluna “Estado”, foi feita a discriminação das variáveis, explicitando as que foram incluídas, excluídas e as que se optou por criar, depois de serem importadas para o ambiente R.

Ir-se-á proceder à justificação da exclusão das variáveis, começando pelo NumAluno, cuja exclusão deve-se ao facto de esta não conter nenhuma informação útil para tirar conclusões, pois todos os números são diferentes e só servem mesmo para identificar o aluno.

Uma outra variável foi o Curso, removida por existirem vários cursos e não se terem conseguido criar padrões ou agrupamentos, de forma que contribuísse o suficiente para a extração de conhecimento, e se considerou que a informação que a mesma continha podia ser mais consolidada. Pelo conseguinte, optou-se por criar uma nova variável a partir desta, a AreaCurso, agrupando os cursos em áreas de Sociologia, Tecnologia e Gestão, de modo a ter uma variável com uma informação mais agregada. Os resultados obtidos dos modelos com a variável AreaCurso foram mais satisfatórios, quando comparados com os obtidos com a variável Curso.

Foi também criada a variável numTestesAno, que representa a quantidade de testes que se realizou por ano.

Em relação à NumTeste, é semelhante à variável descrita na compreensão dos dados, AlunoTeste; a diferença é que esta só diz quantos testes se tem na amostra e quais as perguntas associadas a cada uma, mas a informação de que aluno realizou o teste não se consegue obter a partir desta variável, sendo esta a justificação da exclusão da mesma do modelo, e também porque os resultados obtidos com essa variável mostraram ser inferiores às experiências com a inclusão da mesma. Quanto à NumPergunta, esta variável indica apenas o número de identificação da pergunta, sendo considerada pouco relevante para o estudo, justificando a remoção da mesma.

A exclusão do NumDistrator deve-se ao facto de que, sendo esta e o NumProxima complementares, ou seja, consegue-se obter uma com base na outra, ter-se optado nesta fase escolher uma aleatoriamente para a exclusão, por serem redundantes. Na fase da modelação serão efetuadas experiências sem uma e depois sem a outra, e comparados os resultados obtidos. A que gerou melhor resultado foi incluída no modelo.

Uma outra variável que não foi considerada no modelo foi o Resultado pois é a variável que se pretende estudar. A ResCerta que diz respeito à resposta certa e a ResSelecionada, que diz respeito à resposta selecionada, também foram excluídas, pois estas iriam influenciar os resultados das experiências. Ao se incluir, por exemplo, a resposta certa, esta variável influenciaria o modelo de tal forma que os resultados obtidos seriam muito próximos do 1, o que significa que as variáveis incluídas explicariam o modelo em quase a 100%. Ou seja, o que se pretende é encontrar as variáveis que melhor justifiquem o resultado obtido pelo discente, e ao incluir a resposta certa os resultados que viriam a ser obtidos não seriam elucidativos.

De seguida tem-se a variável CotPergunta, que diz respeito à cotação das perguntas de EM, e como as perguntas todas valiam 1, não se viu a utilidade de inclusão de uma variável que tem apenas um único valor, pois não iria acrescentar nenhum conhecimento ao estudo.

A exclusão da variável Enunciado, e TextoAlínea1,2,3 e 4 deve-se ao facto explicado anteriormente no ponto de compreensão dos dados.

Para exclusão da maioria das variáveis teve-se o mesmo procedimento, primeiro viu-se se os dados relativos à mesma faziam sentido, de seguida refletiu-se na importância das mesmas, salientando que o facto de serem excluídas não quer dizer que não irão ser usadas nas experiências. Muito pelo contrário, na fase seguinte de modelação ir-se-á proceder à realização de várias experiências, comparando os valores com e sem essas variáveis à vez, e daí obter-se-á um leque final de variáveis a serem consideradas para o estudo. Isto porque, até esta fase, a exclusão foi com base na relevância que se acha que cada uma pode ter para o estudo, mas ao analisar os resultados obtidos com e sem essas variáveis ir-se-á saber com auxílio dos modelos de DM as que são relevantes ou não.

3.1.4. CRISP-DM – Modelação

Na fase de **modelação (modeling)** são seleccionadas as técnicas de modelação, é gerado o projeto de teste, e construído e avaliado o modelo. Nesta fase procede-se à execução de várias experiências utilizando as diversas técnicas de DM, implementadas na biblioteca *rminer*. Testou-se várias técnicas, nomeadamente:

- NN
- SVM
- DT
- NB

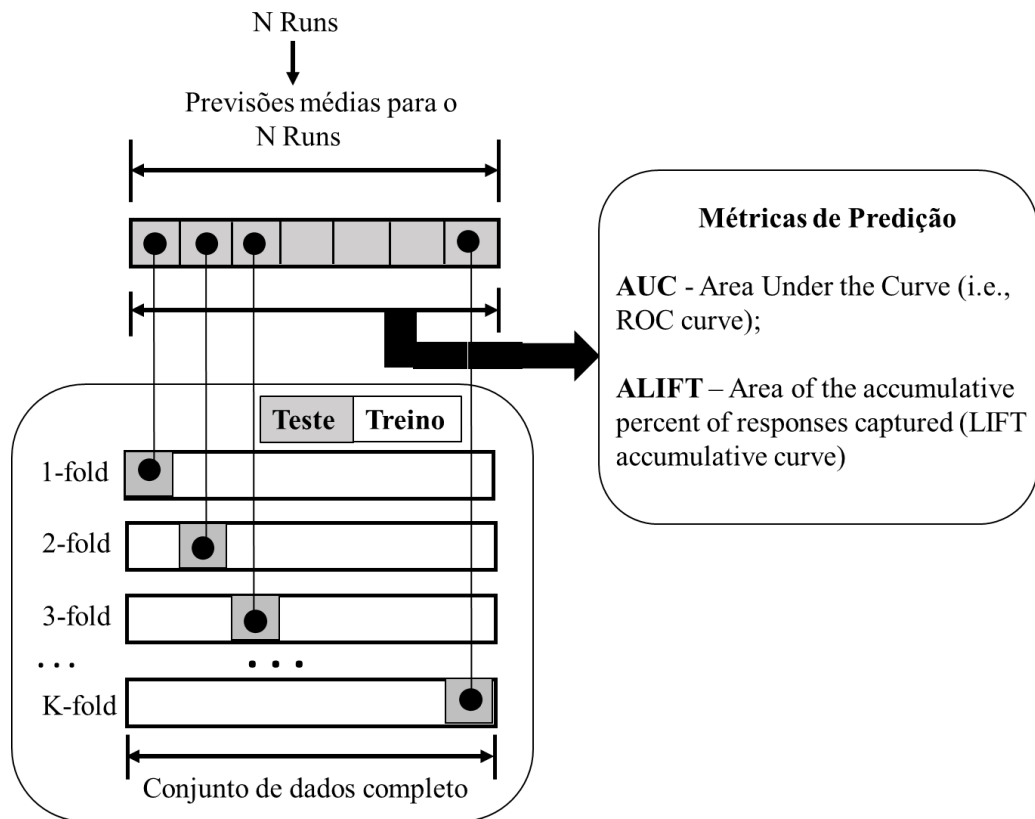


Figura 5 - Processo de modelação

O processo de modelação agrega várias etapas. A figura 5 é uma pequena ilustração de como decorre esse processo. Para testar esses modelos, utilizou-se o procedimento denominado de *k-fold cross-validation*, com $k=10$, de forma a garantir uma validação robusta e conforme é

ilustrado na figura 5, conforme foi também adotado por Moro et al. (2013) e por Moro et al. (2014).

```
### para SVM:
searchSVM1 <- "automatic"
searchSVM2 <- "2^seq(-15,3,4)"
searchSVM3 <- "heuristic5"
### para MLPE:
searchMLPE <- c(0,2,4,6,8)

##experiencias
MLPE <- mining(resultado~., df2, Runs=10,
               model="mlpe", method=c("kfold",10))
savemining(MLPE, "mining_mlpe.output",ascii=TRUE)
SMLPE <- mining(resultado~., df2, Runs=10,
               model="mlpe", search=searchMLPE,
               method=c("kfold",10))
savemining(SMLPE, "mining_smlpe.output",ascii=TRUE)
SVM3 <- mining(resultado~., df2, Runs=10,
               model="svm", search=searchSVM3,
               method=c("kfold",10))
savemining(SVM3, "mining_svm3.output",ascii=TRUE)
SVM1 <- mining(resultado~., df2, Runs=10,
               model="svm", search=searchSVM1,
               method=c("kfold",10))
savemining(SVM1, "mining_svm1.output",ascii=TRUE)
SVM2 <- mining(resultado~., df2, Runs=10,
               model="svm", search=searchSVM2,
               method=c("kfold",10))
savemining(SVM2, "mining_svm2.output",ascii=TRUE)
DT <- mining(resultado~., df2, Runs=10,
             model="dt", method=c("kfold",10))
savemining(DT, "mining_dt.output",ascii=TRUE)
```

```
NB <- mining(resultado~., df2, Runs=10,  
             model="naivebayes", method=c("kfold",10))  
savemining(NB, "mining_nb.output",ascii=TRUE)
```

O excerto de código acima é relativo à função *mining* que pode ser parametrizada para obter o modelo desejado. Esta é bastante relevante, uma vez que executa a técnica referida no parâmetro *model*, obtendo um modelo com base nos dados indicados no parâmetro *method*, e efetuando testes de validação desse modelo. Este processamento pode ser executado tantas vezes quantas as indicadas no parâmetro *Runs*. Com uma seleção aleatória de dados para treino e para teste, é assim possível efetuar uma validação cruzada.

Como se pode observar no código, efetuaram-se experiências para vários modelos e houve alguns modelos como o MLPE (arquitetura de *ensemble de* redes neurais) e o SVM, em que se testaram diferentes valores no parâmetro *search*, sendo o K-fold e o *Runs* sempre 10. Essas são técnicas mais populares e que mostraram melhores resultados, apresentando recursos de aprendizagem que são afetados pela escolha dos seus hiperparâmetros (ex: H, y). Para configurar esses valores, o *rminer* usa a busca em grade e a heurística (utilizada no SVM3), para evitar a sobreposição, (Cortez, 2010).

O resultado é uma estrutura de dados complexa com informação sobre os resultados da execução e como se trata de um processamento moroso e complexo, é possível guardar os resultados para posterior análise, através da função *savemining*, conforme se pode ver no código.

Como foi dito na fase anterior a esta, na modelação seriam efetuadas várias experiências. Testaram-se vários cenários possíveis rodando as variáveis, a fim de considerar apenas as que apresentavam melhores resultados, sendo estas as que têm o estado como “incluídas” na Tabela 1. Para isso usou-se um procedimento do *rminer* de análise de sensibilidade para avaliar a importância do atributo de entrada e caracterizar a influência média na saída alvo (Cortez e Embrechts, 2013).

O procedimento é baseado na medição dos efeitos na saída de um modelo ajustado quando um atributo é variado através do seu domínio de valores e outros atributos são fixados nos seus valores médios. Relativamente à execução nesta fase de modelação, os modelos que mostraram apresentar melhores resultados foram o MLPE e o SVM, sendo esta avaliação feita na fase que se segue.

4. Resultados e Discussão

4.1. Análise das métricas e Variáveis

4.1.1. CRISP-DM - Avaliação

Na fase de **avaliação (evaluation)** são avaliados os resultados e revistos os processos. A avaliação do modelo de classificação ir-se-á basear nas métricas populares AUC e ALIFT, e na curva ROC, cuja interpretação é descrita na secção seguinte.

4.1.2. Resultados preditivos

Para as experiências, foram utilizadas apenas as variáveis cujo estado é incluído na Tabela 1. Foram consideradas 17 variáveis, tendo uma amostra 344 linhas de registos, que contêm informações de respostas dos alunos a cada pergunta de EM. Inicialmente, testaram-se os dados usando a técnica preditiva, que usou 1/3 (dados para teste) dos dados para avaliar a precisão do modelo, calculando as métricas AUC e ALIFT.

A métrica AUC é uma métrica muito comum para problemas de classificação, que apresenta vantagens de ser independente da frequência da classe, tendo o caso em estudo três classes possíveis: “Errado”, “Certo”, e “Não Respondido”. O modelo ideal deve apresentar um AUC de 1,0 enquanto um AUC de 0,5 indica um classificador aleatório. Esse valor indica qual a probabilidade de os dados justificarem o problema em questão, mas como raramente os *inputs* encontram um modelo ideal, o que se procura é efetuar experiências até que se encontre um que esteja mais próximo de 1,0 possível. O mesmo acontece para o valor de ALIFT, que por sua vez representa graficamente a melhoria que um modelo fornece quando comparado com um modelo aleatório e mede a mudança em termos de pontuação de LIFT (TUFFÉRY, 2011) - tenta-se encontrar um valor mais aproximado de 1,0 possível.

Numa fase preliminar, onde foram usados dados destinados para testes, não se utilizaram vários modelos, nem se efetuaram vários testes, pois o que se pretende é saber qual dos modelos testados apresenta melhor performance.

Sendo que os modelos que normalmente apresentam melhores resultados para problemas de classificação são as máquinas de vetores de suporte e redes neurais, estas serão as primeiras a serem testadas nesta fase, salientando que para o segundo modelo irá ser testado segundo a arquitetura MLPE.

Tabela 2 -AUC e ALIFT preditivos

Métricas	Classe	MLPE	SVM
AUC	Certo	0,801	0,774
	Errado	0,790	0,747
	Não respondido	0,833	0,764
ALIFT	Certo	0,622	0,611
	Errado	0,693	0,664
	Não respondido	0,811	0,746

Estas duas primeiras experiências foram sujeitas a uma validação cruzada no sentido de prever melhor o modelo e o desempenho do conjunto de dados, conseguindo assim valores para a AUC bastante acima dos 0,5 de um modelo aleatório, para ambos os modelos, sendo os de MLPE relativamente melhores face aos de SVM.

Analisando os valores, pode-se averiguar que as variáveis escolhidas justificam bem os resultados obtidos nos testes, ou seja, a influência da classificação obtida nos exames deve muito a esses fatores considerados como *inputs*, mesmo sendo esses valores preditivos.

4.1.3. Extração de conhecimento

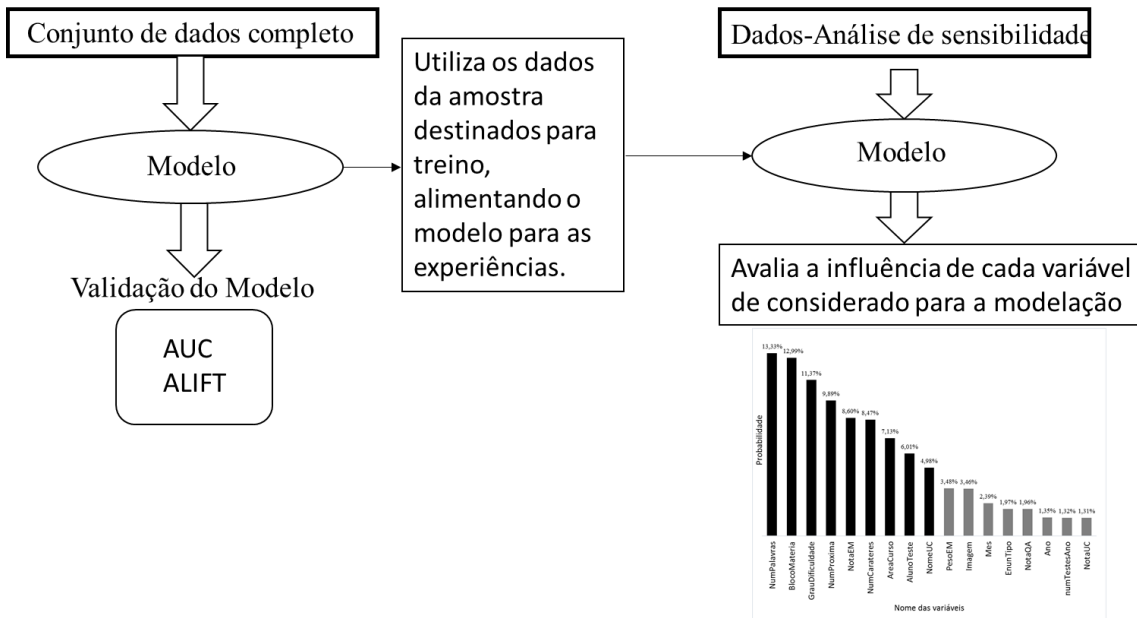


Figura 6 – Processo de Extração de Conhecimento

O esquema acima transcreve como é que se sucede o processo de extração de conhecimento, a partir de experiências usando todo o conjunto de dados para treinar o modelo.

Este procedimento inicia-se com a modelação dos dados, usando todo o conjunto. Com base nesse modelo, é obtido o resultado das métricas de classificação é obtido, a AUC e ALIFT. Conforme se pode ver através da Tabela 3 os resultados são ligeiramente melhores do que os obtidos através do processo *k-fold cross-validation*, conforme esperado, uma vez que estas métricas resultam de verificar para cada instância do conjunto de dados qual o valor modelado, ao passo que o processo de validação anteriormente demonstrado divide os dados em treino e teste, para assegurar uma validação mais robusta.

Tabela 3 - Análise das métricas AUC e ALIFT

	Certo		Errado		Não respondido	
	AUC	ALIFT	AUC	ALIFT	AUC	ALIFT
MLPE	0,826	0,632	0,793	0,695	0,835	0,813
MLPE c/ search	0,828	0,633	0,791	0,694	0,828	0,805
SVM c/ heuristic	0,812	0,626	0,780	0,686	0,807	0,786
DT	0,707	0,583	0,678	0,619	0,717	0,701
NB	0,729	0,592	0,701	0,634	0,757	0,739

O modelo que apresentou melhores resultados foi o MLPE, mostrando um bom desempenho com um AUC consideravelmente bom para ambas as classes. Observando os resultados, pode-se verificar que a taxa AUC é sempre melhor que a de ALIFT, e que o modelo SVM, por sua vez, também apresenta valores consideravelmente bons, muito próximos dos do MLPE.

Um facto interessante que se pode denotar com os resultados é que, para todos os modelos, o “Não respondido” apresenta sempre melhores resultados face aos outros, o que leva a concluir que as variáveis consideradas justificam melhor o facto de o aluno não responder uma pergunta do que modelar o caso do aluno acertar ou errar a mesma. Os modelos DT e NB não apresentam maus resultados, porém são relativamente inferiores ao MLPE e ao SVM, levando à conclusão de que esses modelos não são os melhores para o caso em estudo.

À partida, e a julgar pelos valores das métricas, o modelo MLPE obteve ligeiramente melhores resultados do que o SVM. No entanto, para uma comparação mais clara entre os modelos, pode ser executada a curva ROC recorrendo-se à função *mgraph* da biblioteca *rminer*.

Para uma visualização mais compreensível, será exibida apenas o gráfico das curvas correspondentes à classe Certa, sendo os restantes apresentados no anexo.

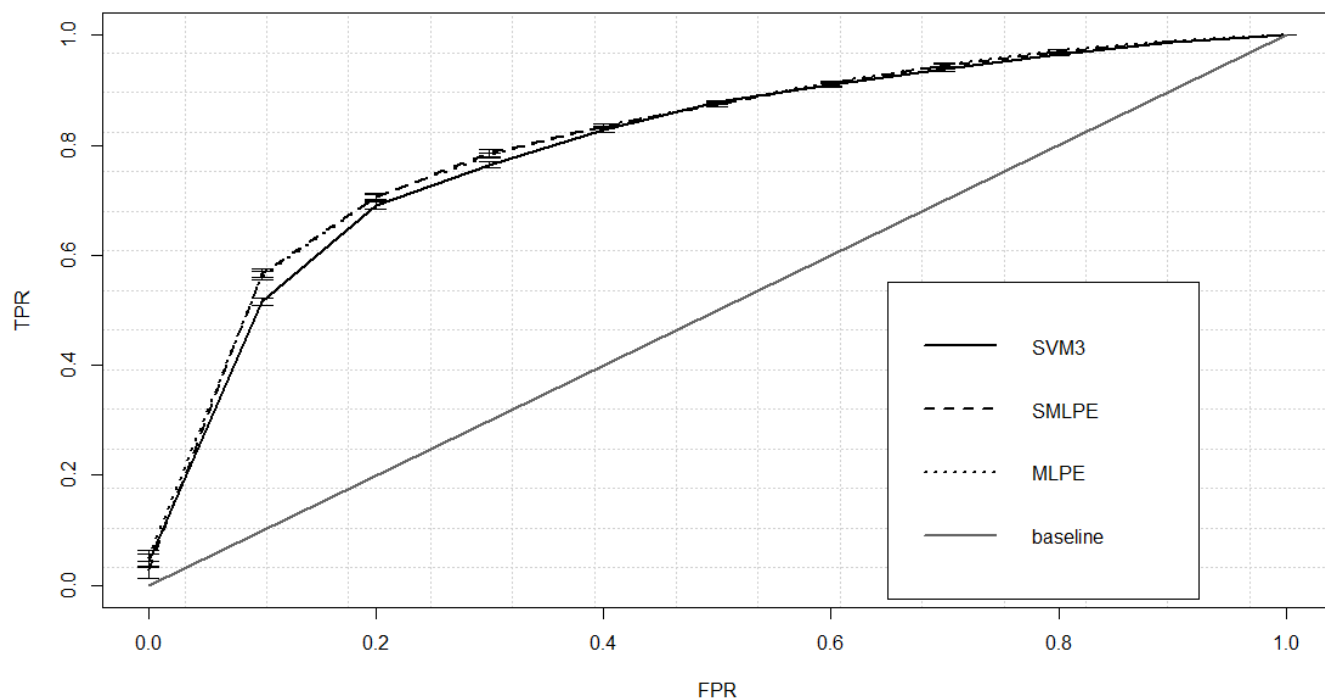


Figura 7 - Curvas ROC, MLPE e SVM

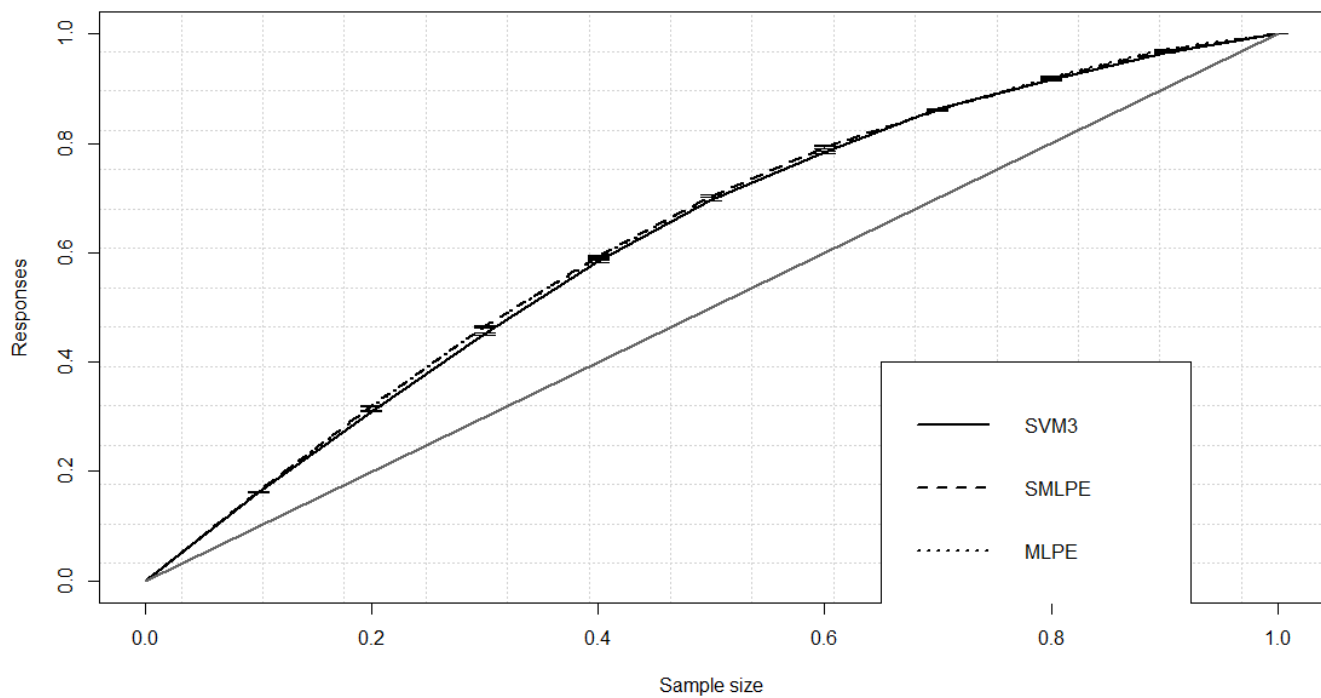


Figura 8 - Curvas ALIFT, MLPE e SVM

Ambos os modelos têm uma capacidade de previsão bastante superior ao modelo aleatório, representado pela reta diagonal que divide a área do gráfico ao meio, podendo ser observado graficamente pelas curvas muito afastadas da reta do modelo aleatório. O MLPE e o SMLPE são os mais afastados, embora estes apresentem valores muito semelhantes, tanto que as curvas dos modelos estão basicamente sobrepostas. Facto que é confirmado pelos valores da métrica AUC, 0,826 e 0,828 para MLPE e MLPE com o parâmetro *search*, respetivamente, com uma diferença de 0,002 entre elas. A curva do modelo SVM, embora próxima das outras duas, está muito próxima da diagonal do modelo aleatório. Complementando esta conclusão com o valor da AUC de 0,812, que não sendo um mau resultado, fica aquém das outras duas.

Já para as curvas LIFT denota-se claramente os resultados não são muito satisfatórios para nenhum dos três modelos, apresentando valores de ALIFT de 0,632, 0,633, 0,626, para MLPE, MLPE com o parâmetro *search* e SVM, respetivamente. Estes valores são bastante inferiores da métrica AUC e muito próximos do modelo aleatório.

4.1.4. Análise das variáveis

Após a análise das métricas, procedeu-se para o estudo das variáveis que foram consideradas para as experiências, e examinaram-se as mais influentes para o problema, realizando-se a análise de sensibilidade das variáveis. Esta análise foi feita utilizando o método *Data-Based sensitivity analysis* (DSA), que seleciona aleatoriamente um subconjunto dos dados usados para treinar o modelo e utiliza-os para fazê-los variar simultaneamente (Cortez e Embrechts, 2013).

A análise de sensibilidade é feita a seguir à fase de modelação, como se vê no esquema da Figura 6. O processo começa por testar as variáveis input, vendo quais as que são mais importantes face às outras, o que significa que o modelo é muito sensível a esta variável e que a mesma é bastante relevante. Aplicando esta análise, obtém-se um output indicando a relevância de cada uma das variáveis consideradas importantes para o modelo. Como se pode observar, há várias variáveis que mesmo sendo consideradas influentes, obtiveram uma percentagem de importância consideravelmente baixa. Deste modo foram selecionadas apenas as variáveis com uma importância acima dos 4% (as barras a preto) para análise.

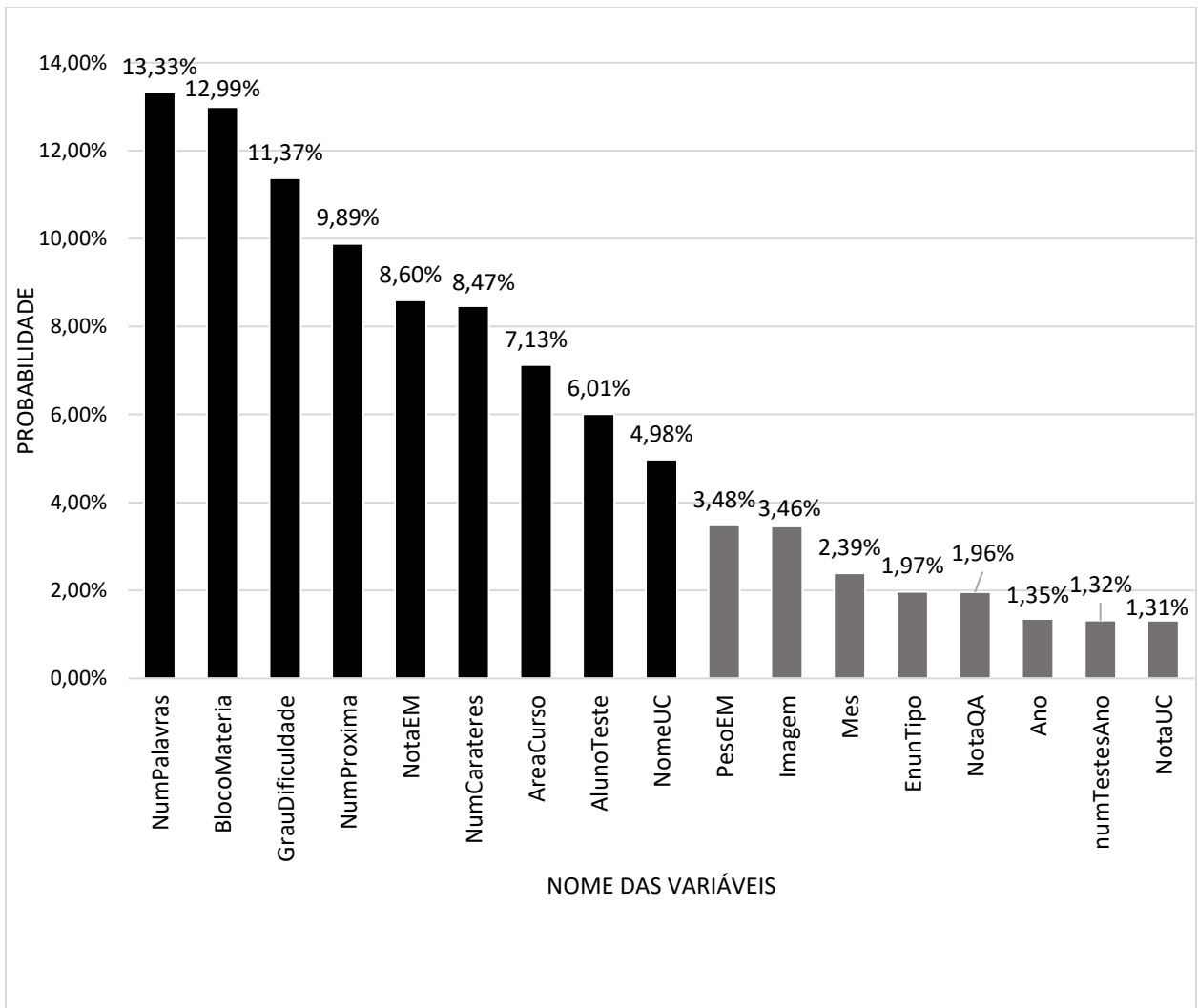


Figura 11 - Relevância das variáveis

Cada uma destas variáveis irá ser analisada, mediante a representação gráfica da probabilidade de acertar ou não responder às perguntas. Optou-se por não se analisar os gráficos que dizem respeito à classe “Errado”, pois esses podem ser inferidos com base na classe “Certo”. Deste modo ir-se-á apenas debruçar sobre os gráficos das variáveis com pelo menos 4% de importância que dizem respeito às classes “Certo” e “Não respondido”.



Figura 12 – Número de Palavras, "Certo"

A quantificação do número de palavras diz respeito ao enunciado, isto é, esta variável indica quão longo é o enunciado da pergunta, lembrando que se trata das perguntas de EM. Como se pode observar pelos gráficos, quanto mais curto for o enunciado maior a probabilidade de o discente acertar a questão, caso contrário a probabilidade de este acertar é bastante reduzida. Uma entrevista feita a alunos que também realizaram testes de questões objetivas do ENEM (Exame Nacional do Ensino Médio) no Brasil, apresentou opiniões dos mesmos a afirmar que o exame foi consideravelmente mais complicado por conter enunciados muito compridos. Uma das alunas inclusive reportou que as provas do ENEM se tornaram cansativas face aos anos anteriores, reclamando que a prova desse ano foi mais difícil. Concordou um outro estudante afirmando: “Demorei mais para raciocinar”. Salientando Santos et al. (2011) que a linguagem e a dimensão do enunciado influenciam a capacidade de uma boa percepção do aluno.

Este fator da má interpretação do enunciado foi considerado por Dubins et al. (2016) como uma possível razão para não asserção da pergunta, dizendo que, o erro do estudante pode surgir por uma variedade de razões, como leitura incorreta de uma questão ou dificuldade de interpretação de uma pergunta mal formulada. Por outro lado, Haladyna et al. (2012) afirmaram que a elaboração de perguntas desta natureza exige cuidados, e que os docentes têm de procurar elaborar questões da maneira mais clara possível, evitando enunciados extensos.

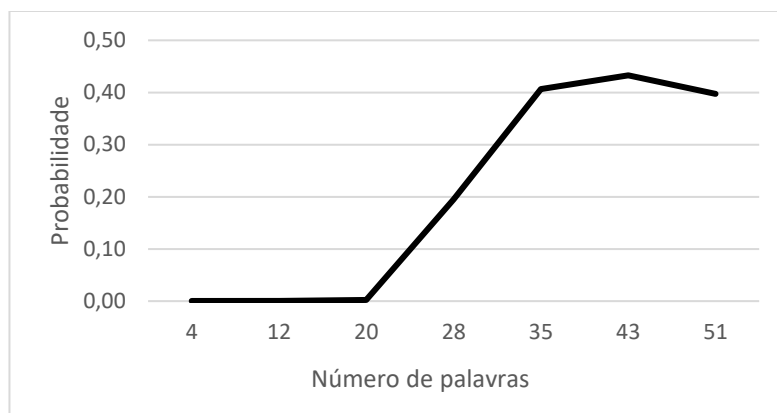


Figura 13 – Número de Palavras, "Não respondido"

Sendo assim, a interpretação do gráfico do “Não Respondido” fica mais fácil de compreender, pois quanto maior for o enunciado maior a probabilidade de este ser mal interpretado pelo discente e, conseqüentemente, de gerar uma insegurança no aluno fazendo com que este não responda à mesma, com medo de errar a pergunta. Se a questão não for bem entendida, ou seja, se não se perceber o que se pretende com a mesma, aumenta o risco de erro, levando a que este não responda ou mesmo que erre a questão em causa, algo que é corroborado pelas figuras 10 e 11.

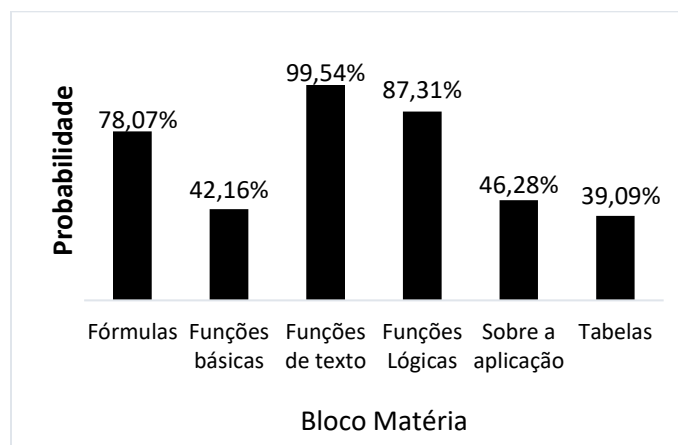


Figura 14 - Bloco de Matéria, "Certo"

Uma outra variável considerada relevante, com aproximadamente 13% de importância, é o Bloco de Matéria, discriminando a probabilidade de acertar ou de não responder, para cada um dos tópicos de matéria considerados nos exames. Como se pode observar, no caso do bloco de

“Funções de texto”, a probabilidade de o aluno acertar a pergunta é muito elevada, tendo também os tópicos de “Funções Lógicas” e “Fórmulas” uma alta chance de acertar, enquanto que se se tratar de questões sobre “Tabelas” a hipótese de acertar é mais reduzida.

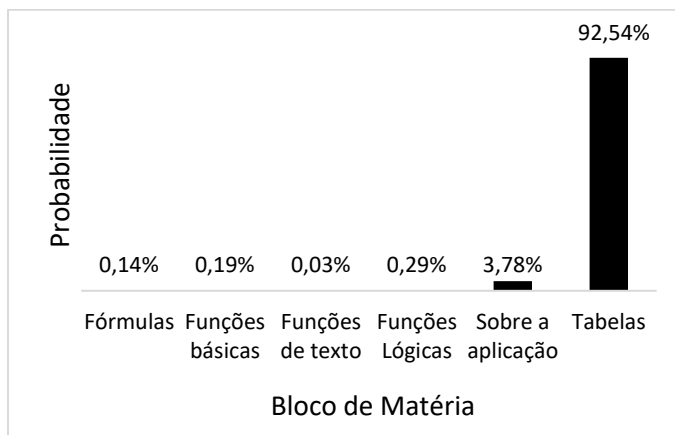


Figura 15 - Bloco de Matéria, "Não respondido"

Analisando o gráfico da Figura 13, os resultados são espectáveis, pois os tópicos em que a probabilidade de acertar é mais alta a hipótese de não responder a estas é muito reduzida, pois se o aluno tem uma alta chance de acertar uma questão não irá deixar a mesma por responder. Tendo assim o bloco “Funções de texto” uma probabilidade 0% de “Não respondida”, e as tabelas em contrapartida de 92,5%, pois na análise do gráfico “Certo” observou-se que a probabilidade de acertar este tópico era inferior a todos os outros considerados para o estudo. Isto deve-se em grande parte ao fato das perguntas relativas às Tabelas serem menos frequentes face às outras, ocupando apenas 3% da amostra, e a desproporcionalidade faz com que a percentagem das “Não respondidas” das Tabelas seja muito mais alta.

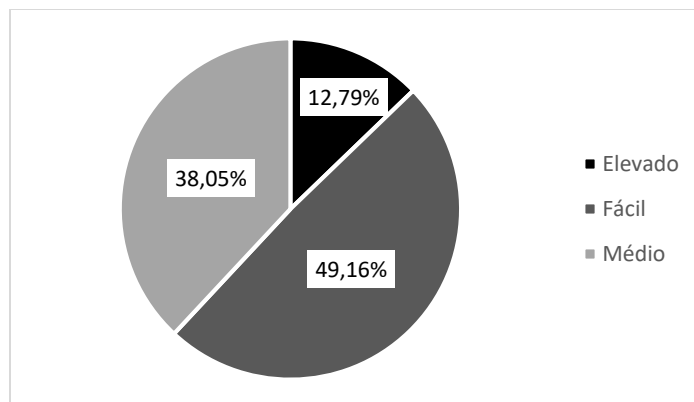


Figura 16 - Grau de Dificuldade, "Certo"

Quanto à terceira variável mais importante, o “Grau dificuldade”, pode ser considerada de fácil justificação, pois sabe-se *a priori* que o aluno terá mais dificuldade em responder uma pergunta considerada difícil e acertar a mesma. Procedendo à interpretação dos gráficos, conforme os graus estipulados para o estudo, nota-se que quando se trata de uma pergunta com um grau muito elevado têm uma probabilidade de sucesso de resposta muito baixa pois exigem mais conhecimento e concentração por parte do aluno. Estas questões normalmente possuem um maior número de fórmulas ou contém funções com condições mais complexas.

Em relação aos graus “fácil” e “médio”, os resultados obtidos são interessantes, pois partindo do senso comum, espera-se que as mais fáceis sejam mais prováveis de serem acertadas do que as médias, o que não acontece segundo os resultados do gráfico. Esse resultado pode derivar da desconfiança do fácil - os alunos por vezes deparam-se com questões que consideram fáceis de mais, e devido a esta desconfiança por vezes acabam por errar por parecer-lhes demasiado óbvia, como explica Andrade et al. (2012). Ressaltando ainda que os principais razões de dificuldade se prendem com relacionar os conhecimentos com informações fornecidas (texto/tabelas/gráficos/esquemas) e a complexidade da matéria.

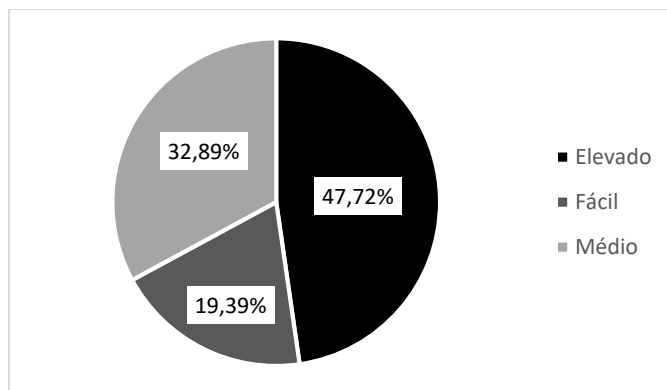


Figura 17 - Grau de Dificuldade, "Não respondido"

Quanto ao resultado “Não respondido” é explicado pela probabilidade mais baixa de acertar uma pergunta difícil, optando-se por não responder ao invés. O mesmo acontece para os graus fácil e médio - é mais provável o aluno acertar, em média, logo a probabilidade de não responder é reduzida.

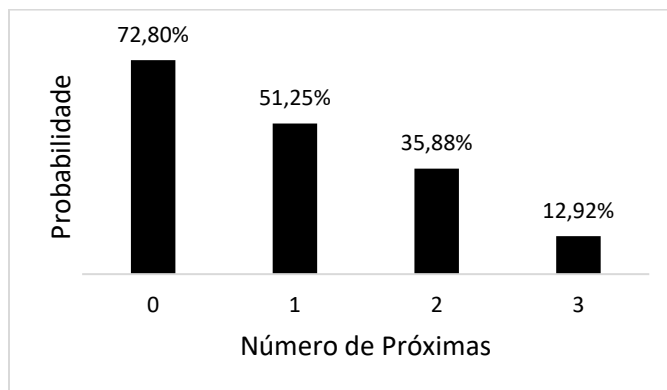


Figura 18 - Número de Próximas, "Certo"

A próxima variável mais influente, com 9,89% de importância, é o NumProxima, que traduz o número de questões próximas da resposta certa existente no leque de opções.

Começando com a interpretação do gráfico da classe “Certo”, é notável que quanto menos opções próximas da opção correta se tiver, maior a probabilidade de o aluno acertar a determinada pergunta, pois quanto mais opções próximas da correta maior a hesitação em arriscar. E nessas situações o que pode acontecer é o aluno ir por exclusão de partes. Dubins et al. (2016) defende que quanto mais opções similares com a resposta correta, maior a probabilidade de acerto por adivinhação, reforçando que, a adivinhação da resposta correta não é

o mesmo que o conhecimento *a priori* da opção correta – o primeiro é probabilístico e o último não o é. Salienta-se que pode haver no máximo três opções próximas, pois as perguntas têm um conjunto com quatro opções sendo uma obrigatoriamente a correta.

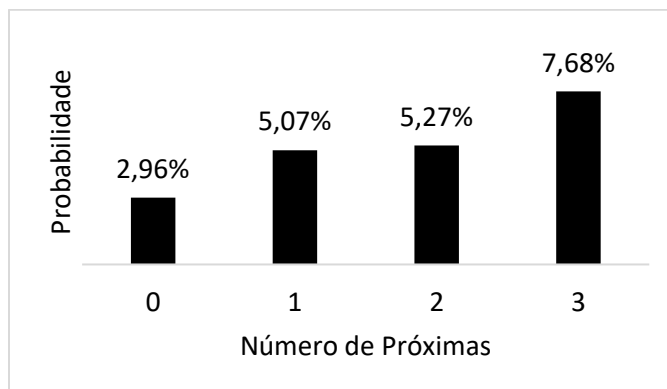


Figura 19 - Número de Próximas, "Não respondido"

Procedendo à análise do gráfico “Não respondido”, as conclusões que se podem tirar são: o aluno quando depara com uma pergunta com nenhuma opção próxima o receio para arriscar é menor, pois nestes casos ele tem menor probabilidade de se enganar escolhendo uma opção que seja próxima da correta, porque esta não existe. Enquanto que se tiver uma opção próxima, o medo de arriscar é maior, porque nesta situação a indecisão pode levar a que o aluno não responda; o mesmo acontece para os casos em que o leque de opções contém duas ou três opções próximas. Nestes casos, o aluno pode ir por exclusão de partes entre as opções ou por adivinhação como já foi dito anteriormente, mas a incerteza irá sempre ser maior, pois terá mais opções que lhe fará questionar a sua decisão, sendo esta uma limitação intrínseca das questões de estilo EM (Dubins et al., 2016).

A relação entre estas duas classes acaba por ser o oposto, pois quanto menos opções próximas tiver maior a chance de o aluno acertar, pois terá menos opções que o deixa intrigado, caso contrário, quanto mais próximas mais indeciso fica o discente, e consequentemente teme errar tendo como consequência a possibilidade de deixar a pergunta em branco.

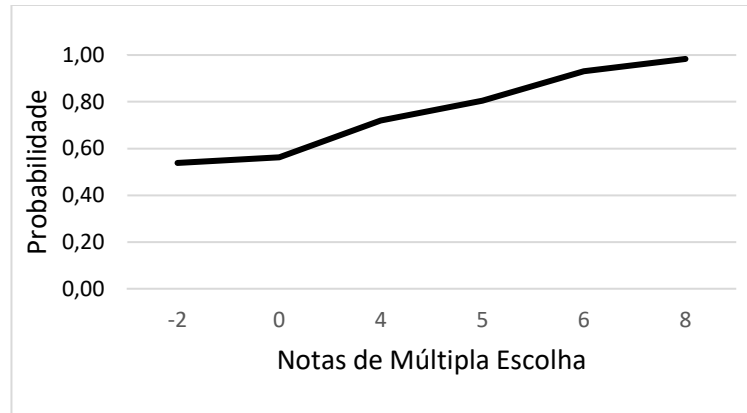


Figura 20 - Notas de EM, "Certo"

A variável “Notas de EM” é a quinta mais importante segundo a análise de sensibilidade feita. Esta indica-nos a nota que o aluno obteve no bloco de EM do teste realizado. A interpretação deste gráfico é trivial, pois é uma variável em que a parte influencia o todo, isto porque a nota que o aluno obtém no bloco de EM é influenciado pelo asserto ou não do discente nas perguntas. No caso do gráfico da classe “Certo”, quanto maior a classificação no bloco maior a probabilidade de acertar, pois a nota obtida em cada questão irá influenciar, como já foi notado, a nota do bloco.

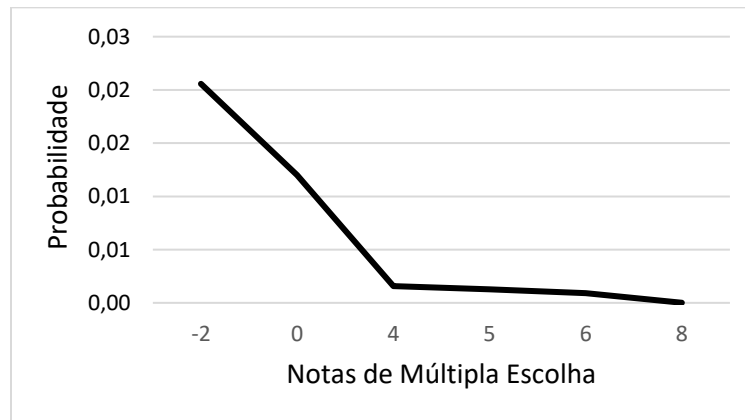


Figura 21 - Notas de EM, "Não respondido"

Em relação à classe “Não respondido”, segue o mesmo raciocínio - quanto menor a classificação, maior a probabilidade de o aluno não responder, ou seja, quanto mais respostas o discente deixar em branco, menor a classificação que este irá obter neste bloco de avaliação. Por

outro lado, segundo o gráfico, quantas mais o aluno responder, maior a probabilidade de obter uma melhor classificação. Ressaltando que essa situação nem sempre é tão linear, pois as perguntas são a descontar - a cada pergunta errada são menos 0,25 valores no bloco, e independentemente do número de questões que esse possa vir a ter, o valor a descontar é sempre o mesmo.

É importante salientar que a categorização das notas presentes no gráfico, são devidas à quantidade de perguntas que cada bloco contém, podendo ser 4, 5, 6 e 8, em que cada pergunta certa vale 1 valor. É possível obter -2 como nota mínima no bloco de EM, uma vez que as perguntas são a descontar.

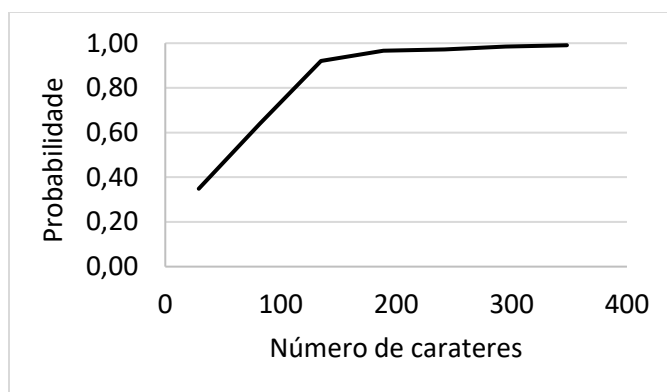


Figura 22 - Número de caracteres, "Certo"

Relativamente à variável NumCarateres, esta é da mesma natureza do número de palavras, pois ambas servem para quantificar o enunciado das perguntas. Porém, são duas formas diferentes de quantificação - uma frase com várias palavras pode não ter vários caracteres pois depende do tamanho das palavras, por isso incluem-se as duas variáveis nas experiências para que se pudesse ter uma quantificação mais precisa.

Esta variável possui 8,47% de importância face às outras. Começando pela análise da classe “Certo”, que indica que quantos mais caracteres se tem num enunciado, ou seja, quanto mais verboso for, maior a probabilidade de o aluno não acertar a mesma, isto porque, como foi apontado na análise do número de palavras, quando se trata de enunciados complexos, a probabilidade de má interpretação por parte dos alunos é maior, implicando uma maior probabilidade que se erre a questão em causa. Por outro lado, quanto mais curto for o enunciado, mais chance o discente tem de acertar a pergunta, pois normalmente quando se depara com

enunciados curtos estes tendem a ser mais objetivos, reforçando Haladyna et al. (2012) que nos enunciados menos verbosos a ideia central da questão é mais explícita e é menos provável que o aluno se confunda.

Analogamente, quando se trata de um enunciado muito longo, a probabilidade de o aluno não responder é bastante elevada, do que se for ao contrário. Isto porque, nos enunciados mais longos, há uma maior chance de uma má interpretação por parte do aluno, e por precaução ou mesmo por receio de arriscar em responder, e ser descontado caso erre a pergunta, este pode preferir em não responder a mesma.

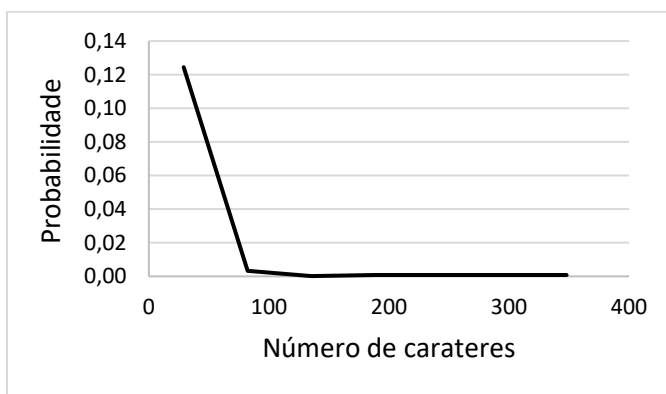


Figura 23 - Número de caracteres, "Não respondido"

Isto é evidente no gráfico “Não respondido” - quanto mais curta a pergunta, menor a probabilidade de o aluno não responder, caso contrário a probabilidade de não responder aumenta significativamente.

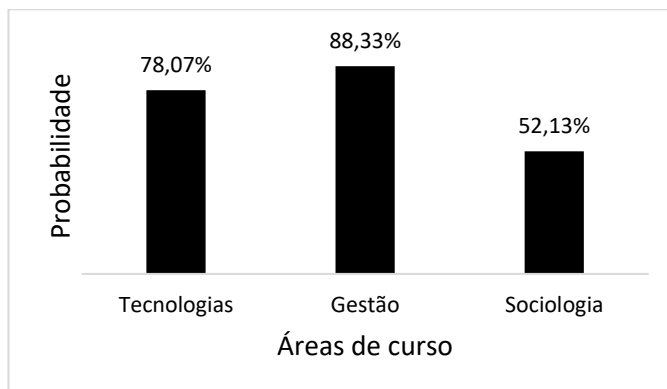


Figura 24 - Áreas de Curso, "Certo"

Passando agora à análise da variável AreaCurso, com 7,13% de importância, verifica-se que a área que tem a maior probabilidade de acertar as questões, ou seja, a área que obtém melhores classificações nos testes, é a de Gestão, estando logo atrás a área de Tecnologia. Porém, esta última tem 17 vezes mais alunos que a de Gestão. É relevante salientar que, conhecendo a amostra mais detalhadamente, a área de Gestão possui uma taxa de probabilidade bastante elevada, pois a maioria dos testes realizados pelos alunos desta área são testes da unidade curricular Introdução a Excel, cujos graus de dificuldade são relativamente mais altos ao de Excel Avançado, que é realizado maioritariamente por alunos da área de Tecnologia.

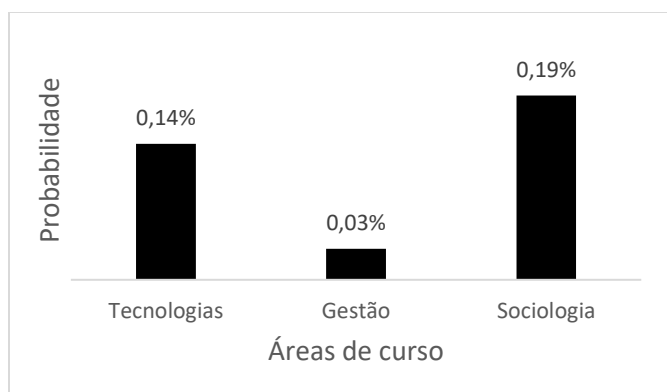


Figura 25 - Áreas de Curso, "Não respondido"

A área cujos alunos obtém uma classificação inferior às outras mencionadas é a de Sociologia, sendo conseqüentemente a área cuja taxa de probabilidade de deixar perguntas sem responder é mais alta.

Enquanto que para as unidades curriculares que constituem a área de Gestão, a probabilidade de realizar um exame e deixar perguntas sem resposta é praticamente nula; já para a área de Tecnologia tem-se ainda alguns casos de questões a serem deixadas em branco.

Mas para ambas as áreas, Tecnologia e Gestão, as taxas de probabilidade de não responder a questões em exames são bastante reduzidas, mesmo para a área de Sociologia, em que a taxa já é ligeiramente maior, porém quase insignificante.

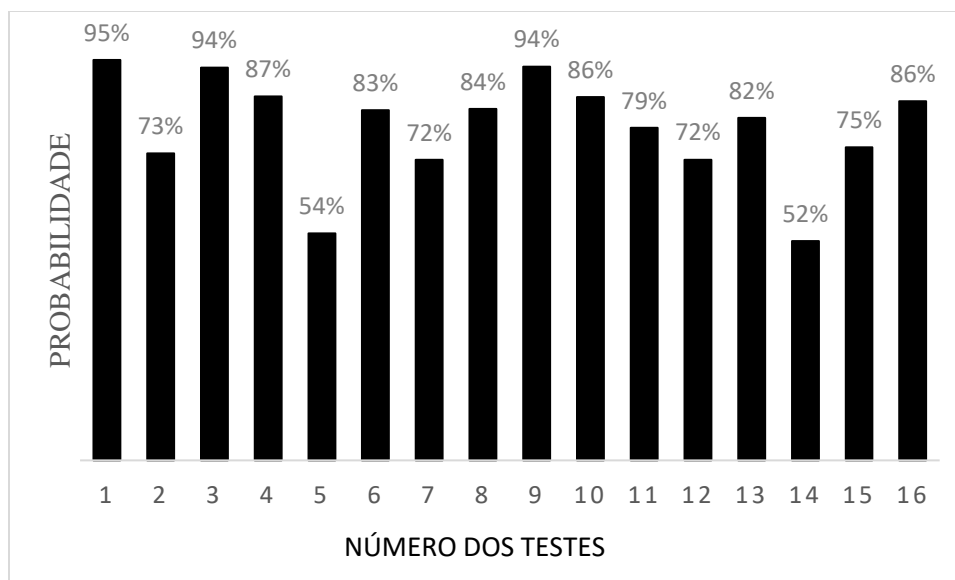


Figura 26 - Número de Testes, "Certo"

A penúltima variável considerada pela análise de sensibilidade efetuada ao leque de variáveis consideradas para o estudo do problema, é o AlunoTeste. Esta variável indica que dos 16 testes que compõe a amostra, qual a probabilidade de o aluno acertar se este realizar o teste 1, 2,3 ... 16. Pela visualização gráfica, consegue-se concluir que os alunos que realizaram o teste 1 tiveram classificações consideravelmente melhores. O que se pode extrair dessa variável que pode vir a ser útil, é que a estrutura dos testes influencia a classificação do discente. Isto porque consegue-se perceber quais os testes em que, normalmente, a classificação é significativamente melhor face às outras, e perceber se a forma como é estruturada o teste, ou os tipos de pergunta que esta contém, tem impacto ou não na classificação.

Os testes que constituem a amostra encontram-se organizados com uma numeração, que não segue nenhuma ordem, mas que facilita identificar o que difere do teste 14 e do teste 1, por exemplo, sendo estes dois em que a probabilidade de acertar as questões é mínima e outra a máxima, respetivamente.

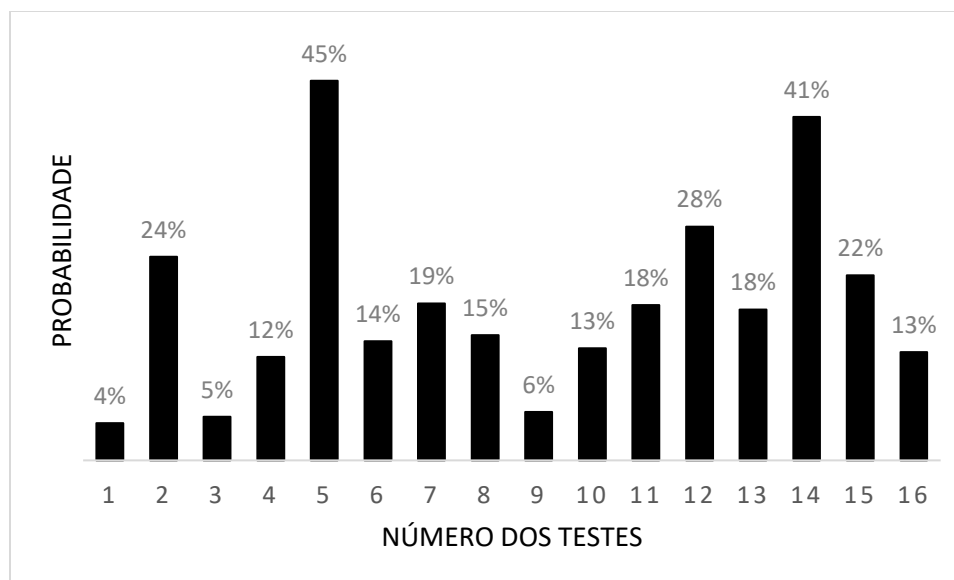


Figura 27 - Número de Testes, "Não respondido"

Como nos outros casos vistos, o resultado para a classe “Não respondido” é um cenário inverso da análise feita anteriormente. Sendo isto natural, pois se os resultados indicam que há testes em que a probabilidade de acertar as perguntas que a constituem é baixa, o espectável é que para esses casos que o aluno não as responda.

Especialmente quando se trata de testes em que se pode ser penalizado por arriscar, no sentido em que responder a uma pergunta em que não se tenha a certeza pode comprometer de forma negativa a classificação final do discente.

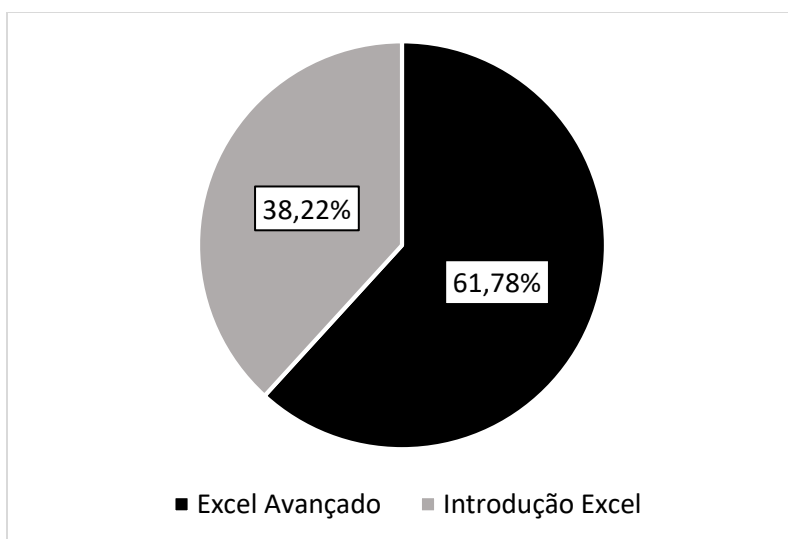


Figura 28 - Unidade Curricular, "Certo"

Por último, tem-se a variável “Nome de unidade curricular”, com apenas 4,98% de importância para o estudo. Esta variável, por sua vez, indica em qual das unidades curriculares é provável obter uma melhor classificação.

Segundo o gráfico da Figura 26, a probabilidade de acertar uma questão de um exame de Excel Avançado é quase o dobro de acertar uma de Introdução a Excel. Esses resultados são devidos aos dados, pois cerca de 95% dos alunos pertencem à unidade curricular Excel Avançado e apenas os 5% restantes são respectivos a alunos de Introdução a Excel.

Esta variável não acrescenta muito conhecimento ao estudo, pois os dados referentes a esta são quase todos iguais, ou seja, não existem padrões que possam vir a ser relacionados, percebendo o porquê da importância da mesma perante as outras variáveis ser tão reduzida.

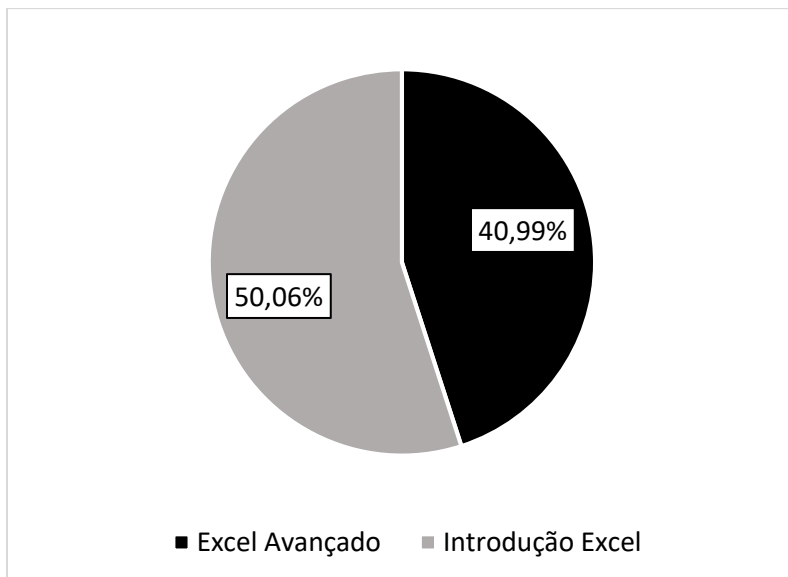


Figura 29 - Unidade Curricular, "Não respondido"

Sendo que a probabilidade de acertar uma questão de um teste de Introdução a Excel é bastante inferior ao comparar com a probabilidade de uma de Excel Avançado, é natural que a probabilidade de não responder uma questão do teste de Excel seja mais alta que a de não responder a uma questão de um exame de Introdução a Excel.

Tendo em conta os resultados acima transcritos, pode-se verificar que cada variável tem a sua particularidade que a difere das restantes, contribuindo para melhor compreensão do que esta na base do problema, ou seja, com esta análise é mais fácil perceber quais das variáveis

consideradas são mais importantes para o estudo e quais os comportamentos das mesmas. Permitem, assim, chegar a conclusões que ajudem a perceber o que leva a que os alunos obtenham uma determinada classificação nos exames de Excel.

4.2. Conclusões

Neste capítulo ir-se-ão apresentar as principais conclusões de todo o trabalho efetuado ao longo desta dissertação, levada a cabo com o intuito de realizar uma análise de uma amostra de testes de Excel disponibilizados pelo ISCTE, a fim de perceber quais das variáveis consideradas para o estudo tem maior influencia no resultado do aluno no bloco de perguntas de EM.

Deste modo, tentou-se durante a realização do trabalho ter sempre os objetivos do estudo como foco, sendo estes:

- Descobrir quais os fatores que mais influenciam as classificações das questões de EM;
- Mostrar quais as principais causas de insucesso nesta tipologia de teste;
- Contribuir para elaboração de questões de EM mais adequadas e com melhor aproveitamento por parte dos discentes.

Foram identificados os principais fatores influentes na classificação que o aluno obtém nas perguntas de EM, dos testes referentes à unidade curricular Excel, que também podem ser consideradas para outras unidades curriculares, cuja estrutura dos exames contenha perguntas de EM.

Uma das principais causas que podem influenciar negativamente os resultados do bloco EM num teste, é o facto de este conter um enunciado muito longo. Isto poderá levar a uma má interpretação por parte do aluno, ou até incompreensão do que se requer da questão, pois sendo o enunciado muito grande a objetividade e a clareza da pergunta é mais reduzida. Por isso o docente, ao elaborar as questões, deverá ter em atenção a clareza do enunciado, pois tratando-se de perguntas de EM, é espectável que o texto seja claro e compreensível, sendo tipos de questões objetivas.

Outro fator muito importante é o bloco de matéria - esta muito específica para o tipo de exames analisados. O docente, com este estudo, terá uma noção de quais os tópicos de matéria que os alunos têm mais dificuldade, e com isto poderá abordá-los de forma mais cautelosa.

O mesmo acontece para o grau de dificuldade da questão - os níveis de dificuldade destes testes foram medidos, tendo em conta o tipo de funções ou a complexidade da fórmula de uma determinada questão. Como já foi tido anteriormente, nestas perguntas em que se exige um esforço de raciocínio maior, tentar alertar os alunos a pontos que os docentes já sabem *a priori* que são distratores. Por exemplo, nos testes do estudo, partes importantes no enunciado eram sempre destacados a negrito para que o aluno tivesse mais atenção a este pormenor.

Deste modo, a chave do sucesso para a elaboração de um bom exame em que se espera obter bons resultados, é saber quais os fatores e de que modo influenciam os resultados. Ou seja, procurar durante a construção, - tarefa do docente e da equipa da determinada unidade curricular -, evitar construir opções com o intuito de confundir o aluno, adversão dada por Haladyna et al. (2012).

Também foi abordada uma taxonomia muito utilizada no seio académico para construção de questões de EM, Taxonomia de Bloom, que indica quais os passos e pontos a ter em consideração na elaboração destes tipos de questões, ao estabelecer níveis de categorização do conhecimento e de como o discente pode alcançar esses níveis de forma progressiva, evitando que o docente exija do aluno algo que ainda não está apto para fazer.

Foi igualmente possível chegar a resultados empíricos que irão ajudar numa melhor projeção de avaliações futuras, sendo o enfoque do mesmo uma área que está sempre a sofrer mudanças, devido a avanços tanto tecnológicos como científicos que cooperam para a evolução no ensino. Educar é uma tarefa difícil, e os docentes tem esta responsabilidade de, enquanto educadores, garantir que o nível de capacidade exigido pelo aluno equivale ao aprendido nas aulas. E para isso é sempre útil e necessário realizar estudos neste âmbito de modo a colaborar para o melhoramento das avaliações realizadas.

4.3. Limitações e Trabalho Futuro

No trabalho exposto foi possível perceber como é que determinados fatores influenciam a classificação obtida no bloco de EM de exames de Excel, e qual a sua na definição de um modelo explicativo, obtido através de técnicas de DM.

A utilidade do modelo explicativo vai no sentido de fornecer conhecimento à equipa docente, que permitirá que estes, na elaboração de questões desta natureza, façam-na de forma adequada, tendo em atenção os fatores considerados importantes neste estudo.

Como foi dito inicialmente, os testes analisados eram constituídos por questões abertas e de EM, porém foram examinados apenas os últimos, e apenas em testes de Excel. Este facto pode ser uma limitação, caso docentes de outras unidades curriculares desejem aproveitar estes resultados para considerar no seu trabalho, pois há fatores que foram considerados pelo modelo que são muito específicos ao problema, ou seja a testes de Excel. Porém, há alguns fatores que podem ser úteis para outras unidades curriculares como o Número de próximas, o Número de palavras e Número de caracteres, pois são comuns a todas em que se elaboram testes com questões de EM.

Este estudo irá servir de ponto de partida para um trabalho em que se vai analisar o outro bloco destes exames, as questões abertas, complementando esta área de investigação e tornando-a mais robusta. Deste modo, a equipa de docentes terá um estudo empírico que os auxiliará na elaboração de exames mais adequados, no sentido em que estes irão ter em consideração alguns aspetos que são fundamentais e mais críticos.

BIBLIOGRAFIA

A

- Amorim, M., Barone, D., & Mansur, A. (2008). *Técnicas de Aprendizado de Máquina Aplicadas na Previsão de Evasão Acadêmica*. Simpósio Brasileiro de Informática na Educação.
- Anderson, W., Krathwohl, D., Airasian, P., Cruikshank, K. A., Mayer, R. E., Pintrich, P., Conklin, J. (2005). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Phi Delta Kappa International.
- Awasthi, N., Bee, S., Mello-e-Souza, D., Carlos, A., & Tinius, D. (2010). *Assessing Accounting Students' Spreadsheet Competency*. Journal of the Academy of Business Education, 11.

B

- Becker, W., & Johnston, C. (1999). *The relationship between multiple choice and essay response question in assessing economics understanding*. Economic Record, 75(4), 348-357.
- Bennett, R., Rock, D., & Wang, M. (1991). *Equivalence of Free-Response and Multiple-Choice Items*. Journal of Educational Measurement, 28(1), 77-92.
- Berson, Alex, and Stephen J. Smith. *Data warehousing, data mining, and OLAP*. McGraw-Hill, Inc., 2001.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The classification of educational goals by a committee of college and university examiners*. New York: Addison-Wesley.
- Buchweitz, B. (1974). *Testes de múltipla escolha e de resposta livre em Física Geral*. Recife.
- Bridgeman, B., Hale, G. A., Lewis, C., Pollack, J., & Wang, M. (1992). Placement validity of a prototype SAT with an essay (Research Rep. 92-28). Princeton, NJ: Educational Testing Service.
- Brito, C. M., Ramos, C. & Carvalho, P., (2006). *Parcerias no Negócio Electrónico*. In: *Parcerias no Negócio Electrónico*. Porto: Sociedade Portuguesa de Inovação, pp. 64-65.
- Baradwaj, B. K., & Pal, S. (2012). *Mining educational data to analyze students' performance*. arXiv preprint arXiv:1201.3417.

C

- Caldeira, C. (2014). *Folhas de Cálculo*. Universidade de Évora
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Conklin, J. (2005). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives Complete*. Educational Horizons.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Cortez, P., & Silva, A. (2008). *USING DATA MINING TO PREDICT SECUNDARY SCHOOL STUDENT PERFORMANCE*. UNIVERSITY OF MINHO, GUIMARÃES, PORTUGAL.
- Cortez, P. Data mining with neural networks and support vector machines using the R/rminer tool. In: *Industrial Conference on Data Mining*. Springer Berlin Heidelberg, 2010. p. 572-583.
- Cortez, P; Embrechts, Mark J. *Using sensitivity analysis and visualization techniques to open black box data mining models*. *Information Sciences*, 2013, 225: 1-17.
- Coutinho, D., & Padilha, T. (2016). *Análise do Desempenho Acadêmico Utilizando Redes Bayesianas: um estudo de caso*. Universidade Federal da Paraíba, Brasil.

D

- Dimitoglou, G., Adams, J. A., & Jim, C. M. (2012). *Comparison of the C4. 5 and a Naïve Bayes classifier for the prediction of lung cancer survivability*. arXiv preprint arXiv:1206.1121.
- Dos Santos Rodrigues, M. A. (2005). *Árvores de Classificação*.
- Drozdek, A. *Estruturas de dados e algoritmos em C++*. São Paulo: Pioneira Thomson Learning, 2005.

F

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Fullan, M., & Hargreaves, A. (2001). *Porque é que vale a pena lutar? O trabalho de equipa na escola*. Porto: Porto Editora.

G

- Gagné, R. M. (1968). *Learning hierarchies*. *Educational Psychologist*, 6, 1-9.

H

- Haladyna, T. M. (2012). *Developing and validating multiple-choice test items*. Routledge.
- Han, J., Kamber, M. & Pei, J., (2011). *Data Mining: Concepts and Techniques*. 3^a ed. Waltham: Morgan Kaufmann.
- Hohensinn, C., & Kubinger, K. (2011). *Applying Item Response Theory Methods to Examine the Impact of Different Response Formats*. *Educational and Psychological Measurement* 71(4) 732–746.

K

- Krathwohl, D. (2002). *A Revision of Bloom's Taxonomy: An Overview*. *Theory Into Practice*, v. 41, n. 4, p. 212-218.

L

- Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). *Data mining techniques and applications—A decade review from 2000 to 2011*. *Expert Systems with Applications*, 39(12), 11303-11311.
- Lissitz, R. W., & Hou, X. (2008). *Multiple choice items and constructed response items: Does it matter?* Obtido em 11 de julho de 2011, de Maryland Assessment Research Center for Education Success.
- Lorena, A. C., & de Carvalho, A. C. (2007). *Uma introdução às support vector machines*. *Revista de Informática Teórica e Aplicada*, 14(2), 43-67.

M

- Maimon, O. & Rokach, L., 2010. *Data Mining and Knowledge Discovery Handbook*. 2^a ed. London: Springer.
- Martinez, M. E. (1991). *A comparison of multiple-choice and constructed response figural response items*. *Journal of Educational Measurement*, 28, 131-145.
- Mendes, A. B. Metodologias de Data Mining. Ok, computer!: confluências na informática., 2011, 29-30.
- Miller, H., Williams, R. G., e Haladyna, T. M. (1978). *Beyond facts: Objective ways to measure thinking*. Englewood Cliffs, NJ: Educational Technology.
- Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: the case of Las Vegas Strip. *Tourism Management Perspectives*, 23, 41-52.
- Moro, Sérgio; CORTEZ, Paulo; LAUREANO, Raul. A data mining approach for bank telemarketing using the rminer package and r tool. 2013.
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22-31.

N

- National Assessment of Educational Progress (1988). *Science objectives: 1990 alicaslarak*. Princeton, NJ: Educational Testing Service.
- Nogueira, D. R. P. (2014). Agile Data Mining: uma metodologia ágil para o desenvolvimento de projetos de data mining.

P

- Pacheco, J. A. (2002). Critérios de avaliação na escola.
- Pinto, A. C. (2001). *Factores relevantes na avaliação escolar por perguntas de escolha múltipla*. *Psicologia, Educação e Cultura*, 5 (1), 23-44.

R

Raisinghani, M., (2004). *Business Intelligence in the Digital Economy: Opportunities, Limitations and Risks*. s.l.:Idea Group Publishing.

Rodriguez, M. C. (2003). *Construct Equivalence of Multiple-Choice and Constructed-Response Items: A Random Effects Synthesis of Correlations*. *Journal of Educational Measurement*, 40(2), 163-184.

Royer, J. M., Cisero, C. A., e Carlo, M. S. (1993). *Techniques and procedures for assessing cognitive skills*. *Review of Educational Research*, 63, 201-243.

S

Sheu, M. Z., & Wang-Chan, W. (2006). *A knowledge assimilation schema for acquiring technical knowledge*. *Journal of Information Systems Education*, 17(2), 223.

Santos, J. R. D. (2011). A constituição do enunciado nas provas do ENEM e do ENADE: uma análise dos aspectos semiológicos da relação língua-imagem sob a ótica dos estudos do discurso.

T

Traub, R. E. (1993). *On the equivalence of the traits assessed by multiple-choice and constructed-response tests. Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*, 29-44.

Treadwell, G., Estep, M., Smith, K. D., & Merritt, K. L. (2013). *Spreadsheet Proficiency in Business School Students: A Preliminary Study of Student Job Preparedness*. *Association of Business Information Systems (ABIS) Refereed Proceedings*, 87-100.

Turban, E., Sharda, R. and Delen, D. (2010). *Decision Support and Business Intelligence Systems – 9 th edition*, Prentice Hall Press, USA.

Tufféry, S. (2011). Association Analysis. *Data Mining and Statistics for Decision Making*, 287-299.

V

Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). *Predicting academic performance by data mining methods*. *Education Economics*, 15(4), 405-419.

W

Walstad, W. B., & Becker, W. E. (1994). *Achievement differences on multiple-choice and essay tests in economics*. *American Economic Review, Papers and Processing*, 84(2), 193-196.

Wirth, R. (2000). Crisp-dm position statement. In 6th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, USA.

ANEXO

Taxonomia de Bloom

Tabela 4 - Taxonomia Original, Fonte: Bloom et al.(1956), Bloom (1986), Driscoll (2000) e Krathwohl (2002).

Categoria	Palavras-chave (verbos), e tecnologias para aprendizagem (atividades)
1.Conhecimento: Lembrar informações e conteúdos.	Palavras-chave: organiza, define, descreve, identifica, reconhece, reproduz, seleciona, relacione, enumera, lista, aponte, aponta, combina. Tecnologias: leitura, pesquisa na internet, memorização.
2.Compreensão: Compreender e dar significado ao conteúdo compreendido em forma escrita, oral, diagrama, etc.	Palavras-chave: compreende, converte, defende, distingue, estima, explica, generaliza, dá um exemplo, infere, interpreta, parafraseia, reescreve, resume, ilustre, resolve, produza. Tecnologias: criar uma analogia, participar na aprendizagem cooperativa, tomar notas, contar histórias.
3.Aplicação: Usar informações, métodos e conteúdos aprendidos em novas situações.	Palavras-chave: aplica, mudanças, computa, constrói, demonstra, descobre, manipula, modifica, opera, prediz, prepara, produz, relaciona, mostra, resolve, usa, esboce. Tecnologias: aprendizagem colaborativa, criar um processo, blog, prática.
4.Análise: Separa material ou conceitos em partes componentes para que sua estrutura organizacional possa ser compreendida.	Palavras-chave: analisa, divide, compara, contrasta, diagrama, desconstrói, diferencia, discrimina, distingue, identifica, ilustra, infere, contornos, relata, seleciona, separa Tecnologias: debater, questionar o que aconteceu, executar um teste.
5.Síntese: Constrói uma estrutura ou padrão a partir de	Palavras-chave: categoriza, combina, compila, compõe, cria, concebe, projeta, explica, gera, modifica, organiza, planeja, rearranja, reconstrói, relata, reorganiza, revê, reescreve, resume, conta, escreve.

diversos elementos.	Tecnologias: ensaio, trabalho em rede.
6.Avaliação: Faça julgamentos sobre o valor de ideias ou materiais.	Palavras-chave: avalia, compara, conclui, contrasta, critica, critica, defende, descreve, discrimina, avalia, explica, interpreta, justifica, relaciona, resume Tecnologias: pesquisa, blogs.

Tabela 5 - Taxonomia Revisto, Fonte: Anderson, Krathwohl e Airasian (2001)

Categoria	Palavras-chave (verbos), e tecnologias para aprendizagem (atividades)
1.Lembrar: Recuperar informações aprendidas anteriormente.	Palavras-chave: define, descreve, identifica, sabe, listas, nomes, contornos, recordações, reconhece, reproduz, seleciona. Tecnologias: leitura, marcação de livros, memorização.
2.Entender: estabelecer uma conexão entre algo novo e o conhecimento previamente adquirido, reproduzindo a informação com suas “próprias palavras”.	Palavras-chave: compreende, converte, defende, distingue, estima, explica, estende, generaliza, dá um exemplo, infere, interpreta, parafraseia, prediz, reescreve, resume. Tecnologias: criar uma analogia, participar na aprendizagem cooperativa, tomar notas, contar histórias, pesquisa na Internet.
3.Aplicar: executar ou usar um procedimento numa situação específica e pode também abordar a aplicação de um conhecimento numa situação nova.	Palavras-chave: aplica, altera, computa, constrói, demonstra, descobre, manipula, modifica, opera, prevê, prepara, produz, relata, mostra, resolve, usa. Tecnologias: aprendizagem colaborativa, criar um processo, blog, prática.
4.Analisar: dividir a informação em partes relevantes e irrelevantes, importantes e menos importantes e entender a	Palavras-chave: analisa, divide, compara, contrasta, diagrama, desconstrói, diferencia, discrimina, distingue, identifica, ilustra, infere, contornos, relata, seleciona, separa. Tecnologias: debater, questionar o que aconteceu, executar um teste.

inter-relação existente entre as partes.	
5.Avaliar: realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de eficiência e eficácia	Palavras-chave: avalia, compara, conclui, contrasta, critica, critica, defende, descreve, discrimina, avalia, explica, interpreta, justifica, relaciona, resume. Tecnologias: pesquisa, blogs.
6.Criar: colocar elementos junto com o objetivo de criar uma nova visão, uma nova solução, estrutura ou modelo utilizando conhecimentos e habilidades previamente adquiridos.	Palavras-chave: categoriza, combina, compila, compõe, cria, concebe, projeta, explica, gera, modifica, organiza, planeja, rearranja, reconstrói, relata, reorganiza, revê, reescreve, resume, conta, escreve. Tecnologias: Criar um novo modelo, escrever um ensaio, trabalho em rede.

Tabela 6 - Matriz Bidimensional, Anderson et al. (2011) ;Clark, Chopeta (2004); Clark, Mayer (2007)

Dimensão Conhecimento	Dimensão processo cognitivo					
	Lembar	Entender	Aplicar	Analisar	Avaliar	Cirar
Fatos	Objetivo1					
Conceitos		Objetivo2	Objetivo2			
Processos						
Procedimentos						
Princípios					Objetivo3	
Metacognitivo						