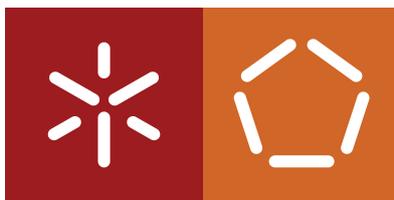


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Jéssica Andreia Fernandes Lemos

**Análise e Avaliação de Desempenho de
Alunos num Sistema de eLearning**

June 2022



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Jéssica Andreia Fernandes Lemos

**Análise e Avaliação de Desempenho de
Alunos num Sistema de eLearning**

Dissertação de Mestrado
Mestrado Integrado em Engenharia Informática

Dissertação supervisionada por
Professor Orlando Manuel de Oliveira Belo

June 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

LICENÇA CONCEDIDA AOS UTILIZADORES DESTE TRABALHO:



CC BY

<https://creativecommons.org/licenses/by/4.0/>

AGRADECIMENTOS

Terminada a presente dissertação e, por consequência, o meu Mestrado Integrado em Engenharia Informática, não poderia deixar de agradecer a todos aqueles que contribuíram ao longo deste percurso.

Começo por agradecer ao meu orientador, Professor Orlando Belo, pelo seu acompanhamento e disponibilidade ao longo do último ano.

De seguida, a toda a minha família, em especial à minha mãe, ao meu pai e irmão, que são os meus maiores pilares. Agradeço por me apoiarem incondicionalmente desde o primeiro dia, por acreditarem nas minhas capacidades e darem-me toda a força necessária para ultrapassar os obstáculos e seguir os meus sonhos.

Por fim, quero agradecer aos meus amigos por me terem acompanhado ao longo do meu percurso académico, marcado pela partilha de conhecimento e entreaajuda. Serei eternamente grata pela vossa amizade e pelo que alcançamos juntos, levo amigos para a vida.

Estes anos foram os melhores e mais importantes da minha vida, marcados pelo crescimento e evolução, tanto pessoal como profissional, por isso a todos quero manifestar os meus sinceros agradecimentos.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

ABSTRACT

The evolution of technology has led to the appearance of some gaps in several sectors. In the area of education, one of the problematic issues pointed out by teachers is the lack of more specific and detailed data on the students' learning process, besides the visualisation of the final assessment results. Due to the scarcity of information about the process their students follow to achieve a given grade, teachers have some difficulty in acting effectively and continuously in this process of learning and teaching their students. This is due to the fact that teachers have no way of identifying what knowledge students have already acquired and what they still need to acquire in order to improve. Such circumstance may contribute to the students' failure. Being this an age characterized by the evolution of technology, in which many problems are solved with it, it is usual to point out as a possible solution to these problems the adaptation of the education sector to new technologies, since in that way it would be possible to fill some of the gaps using tools already developed. One of the tools that emerged to solve the various problems associated with the monitoring and evaluation of student performance were the eLearning systems, in particular its aspect of results analysis through specialized dashboards, since they allow its realization practically in real time. The importance of restructuring the educational sector, through the insertion of new technological tools, is, therefore, unavoidable. The needs of students and teachers alike are changing and evolving, and to cope with this we need versatile and flexible tools like analytics systems and their dashboards.

KEYWORDS Multidimensional Data Systems, eLearning Systems, Performance Analysis Systems, Data Mining, Data Visualization, Dashboards.

RESUMO

A evolução da tecnologia conduziu ao aparecimento de algumas lacunas em diversos sectores. Na área da educação, uma das questões problemáticas apontadas pelos professores é a falta de dados mais específicos e detalhados sobre o processo da aprendizagem dos alunos, para além da visualização dos resultados das avaliações finais. Devido à escassez de informação sobre o processo que os seus estudantes seguem para atingirem uma dada classificação, os professores têm alguma dificuldade em atuar de forma eficaz e contínua neste processo de aprendizagem e de ensino dos seus estudantes. Tal deve-se ao facto dos professores não terem como identificar qual o conhecimento já adquirido pelos alunos e aquele que ainda precisam de adquirir para melhorar. Tal circunstância pode contribuir para o insucesso dos alunos. Sendo esta uma Era caracterizada pela evolução da tecnologia, na qual muitos dos problemas são solucionados com ela, é usual apontar-se como uma possível solução para estes problemas a adaptação do setor da educação às novas tecnologias, uma vez que dessa forma seria possível colmatar algumas das lacunas recorrendo-se a ferramentas já desenvolvidas. Uma das ferramentas que surgiu para solucionar os vários problemas associados ao acompanhamento e avaliação do desempenho dos alunos foram os sistemas de *eLearning*, em particular a sua vertente de análise de resultados através de *dashboards* especializados, uma vez que estes permitem a sua realização praticamente em tempo real. A importância de reestruturar o setor educacional, através da inserção de novas ferramentas tecnológicas, é, assim, algo incontornável. As necessidades dos estudantes como dos professores também estão a mudar e a evoluir e, para lidar com isso, precisamos de ferramentas versáteis e flexíveis como os sistemas de análise e os seus *dashboards*.

PALAVRAS-CHAVE Sistemas Multidimensionais de Dados, Sistemas de *eLearning*, Sistemas de Análise de Desempenho, Mineração de Dados, Visualização de dados, Dashboards.

CONTEÚDO

| | | |
|-------|--|----|
| 1 | INTRODUÇÃO | 3 |
| 1.1 | Contextualização | 3 |
| 1.2 | Motivação e Objetivos | 4 |
| 1.3 | Trabalho Realizado | 5 |
| 1.4 | Organização da Dissertação | 5 |
| 2 | SISTEMAS DE AVALIAÇÃO | 7 |
| 2.1 | Uma Breve Introdução | 7 |
| 2.2 | Sistemas Existentes | 7 |
| 2.2.1 | Moodle | 7 |
| 2.2.2 | Blackboard | 8 |
| 2.2.3 | Claroline | 9 |
| 2.2.4 | ATutor | 10 |
| 2.2.5 | Dokeos | 11 |
| 2.2.6 | Docebo | 12 |
| 2.3 | Análise Comparativa | 13 |
| 3 | ANÁLISE E AVALIAÇÃO DO DESEMPENHO DOS ESTUDANTES | 15 |
| 3.1 | Contextualização e Objetivos | 15 |
| 3.2 | Modelos de Avaliação do Desempenho de Estudantes | 16 |
| 3.3 | Análise de Modelos de Avaliação | 19 |
| 3.4 | Análise Comparativa | 21 |
| 4 | UM SISTEMA PARA A AVALIAÇÃO DE DESEMPENHO | 23 |
| 4.1 | Caracterização do Modelo de Avaliação | 23 |
| 4.2 | Angariação e Pré-processamento de Dados | 25 |
| 4.3 | Classificação de Desempenho | 26 |
| 4.4 | Previsão de Desempenho | 32 |
| 4.5 | Clustering de Desempenho | 33 |
| 5 | ANÁLISE DOS MODELOS IMPLEMENTADOS | 37 |
| 5.1 | Classificação do Desempenho | 37 |
| 5.2 | Previsão | 41 |
| 5.3 | Clustering do Desempenho | 43 |
| 6 | VISUALIZAÇÃO E ANÁLISE DE DADOS DE DESEMPENHO | 46 |
| 6.1 | Dashboards | 46 |
| 6.2 | Técnicas e Componentes de Visualização | 47 |

| | | |
|-----|--|----|
| 6.3 | Modelos e Componentes Implementados | 49 |
| 6.4 | Caracterização e Funcionamento dos <i>Dashboards</i> | 50 |
| 7 | CONCLUSÕES E TRABALHO FUTURO | 54 |
| 7.1 | Conclusões | 54 |
| 7.2 | Trabalho Futuro | 56 |

LISTA DE ACRÓNIMOS

- AI** *Artificial Intelligence*. 16
- CMS** *Course Management System*. 7
- CRISP-DM** *CRoss Industry Standard Process for Data Mining*. vi, 5, 23, 24, 54
- CV** *cross validation*. 19, 32, 38, 39, 41, 55
- DM** *Data Mining*. vi, 4–6, 15–18, 20–24, 29, 54, 56
- DT** *Decision Tree*. vi, viii, 19–22, 24, 27–31, 37–41, 49, 55
- EDM** *Education Data Mining*. vi, 16–21
- HPO** *Hyperparameter Optimization*. 30
- LA** *Learning Analytics*. 17
- LMS** *Learning Management System*. viii, 5, 7, 9, 11–14
- LSA** *Latent Semantic Analysis*. 19
- LSTM** *Long Short-term Memory*. 56
- ML** *Machine Learning*. 17
- MLP** *Multilayer Perceptron*. 20
- NN** *Neural Network*. 20, 27, 29, 56
- RF** *Random Forest*. vi, viii, 24, 28–31, 39–42, 49, 50, 55
- RNN** *Recurrent Neural Network*. 56
- SVM** *Support Vector Machines*. vi, viii, 24, 27, 29–31, 40, 41, 55
- VLE** *Virtual Learning Environment*. 7

LISTA DE FIGURAS

| | | |
|-----------|--|----|
| Figura 1 | O ambiente de trabalho do Moodle | 8 |
| Figura 2 | O ambiente de trabalho do Blackboard | 9 |
| Figura 3 | O ambiente de trabalho do Claroline | 10 |
| Figura 4 | O ambiente de trabalho do ATutor | 11 |
| Figura 5 | O ambiente de trabalho do Dokeos | 12 |
| Figura 6 | O ambiente de trabalho do Docebo | 13 |
| Figura 7 | Descoberta de conhecimento educacional e processo de DM – figura adaptada de (Romero e Ventura, 2013) | 17 |
| Figura 8 | Principais áreas relacionadas com a EDM – figura adaptada de (Romero e Ventura, 2013) | 18 |
| Figura 9 | Mineração de dados educacionais – figura adaptada de (Romero e Ventura, 2007) | 18 |
| Figura 10 | Processo de mineração de dados | 19 |
| Figura 11 | Modelo CRISP-DM – figura adaptada de (Wirth, 2000) | 24 |
| Figura 12 | Estrutura de uma DT | 28 |
| Figura 13 | Estrutura de uma RF | 29 |
| Figura 14 | Processo do classificador SVM | 29 |
| Figura 15 | Comparação <i>grid search</i> e <i>random search</i> – figura adaptada de (Bergstra e Bengio, 2012) | 31 |
| Figura 16 | Ilustração do processo de <i>clustering</i> – figura adaptada de (Xu e Li, 2005) | 33 |
| Figura 17 | Identificação do <i>Elbow</i> para o processo <i>k-prototypes</i> | 35 |
| Figura 18 | Definição do <i>Elbow</i> para o processo <i>k-means</i> | 36 |
| Figura 19 | Resultado obtido: Árvore resultante da aplicação do modelo DT | 38 |
| Figura 20 | Resultado obtido: Árvore resultante da aplicação do modelo RF - 1º estimador | 40 |
| Figura 21 | Previsão do desempenho (taxa de acerto) de um aluno | 42 |
| Figura 22 | Previsão do desempenho (taxa de acerto) por subdomínio | 42 |
| Figura 23 | Resultados obtidos: <i>clustering</i> com <i>k-prototypes</i> | 43 |
| Figura 24 | Resultados obtidos: <i>clustering</i> com <i>k-prototypes</i> | 44 |
| Figura 25 | Resultados obtidos: <i>clustering</i> com <i>k-means</i> | 45 |
| Figura 26 | Diagrama de escolha de tipo de visualização – figura adaptada de (Tim, 2019) | 48 |
| Figura 27 | <i>Dashboard</i> de classificação | 51 |

| | | |
|-----------|---|----|
| Figura 28 | <i>Dashboard</i> de previsão | 52 |
| Figura 29 | <i>Dashboard</i> de <i>clustering</i> | 53 |

LISTA DE TABELAS

| | | |
|----------|--|----|
| Tabela 1 | Análise comparativa de LMS | 14 |
| Tabela 2 | Análise comparativa dos modelos estudados | 22 |
| Tabela 3 | Resultado da <i>feature selection</i> | 26 |
| Tabela 4 | Dados em análise no processo de <i>clustering</i> | 35 |
| Tabela 5 | Síntese dos resultados obtidos com o modelo DT | 39 |
| Tabela 6 | Síntese das métricas obtidas com o modelo RF | 40 |
| Tabela 7 | Síntese dos resultados obtidos com o modelo SVM | 41 |

INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Nos dias de hoje, podemos constatar algumas mudanças rápidas e desenvolvimentos notórios em algumas tecnologias. Esta evolução na área da tecnologia, com o aparecimento e popularização da Internet bem como a consolidação dos dispositivos móveis, conduziu ao aparecimento de algumas lacunas. As inúmeras situações de excesso de informação, associadas à crescente pressão para acelerar o processo de tomada de decisão, prejudicam seriamente a eficiência e qualidade do processo de tomada de decisão. De modo a atenuar estes problemas, que emergiram com os avanços tecnológicos, tanto os indivíduos como as instituições foram forçados a tomarem consciência das suas necessidades e adaptar-se à Era atual. Uma das áreas que está suscetível a mudanças é a educação.

Um dos principais objetivos de qualquer sistema educativo é dotar os estudantes com o conhecimento e as aptidões necessárias para alcançarem carreiras bem sucedidas dentro de um período específico. A eficácia com que os sistemas educacionais globais cumprem este objetivo é um fator determinante tanto do progresso económico como social (Algarni, 2016). Contudo, os professores, de momento, deparam-se com dificuldades em atuar de forma eficaz na aprendizagem e ensino dos seus estudantes, uma vez que estes apenas têm acesso aos resultados das avaliações finais (Maldonado et al., 2012). Consequentemente, não é possível identificar alguns dos problemas e dificuldades que os alunos têm durante o processo, o que é apontado como um dos fatores que contribuem para a elevada percentagem de insucesso e desistência académica.

Atualmente, considera-se que a avaliação do desempenho dos alunos não deve ser vista como algo separado do processo de aprendizagem, uma vez que esta é uma parte contínua e integrante de todo o processo. Ao revelar o conhecimento já adquirido pelos alunos e aquele que estes ainda precisam de aprender e melhorar, permite-se que os professores identifiquem e analisem o processo de ensino e aprendizagem e, desta forma, desenvolver uma estrutura de aprendizagem melhorada e adequada para cada um dos seus alunos. Informação oportuna e específica fornecida pela avaliação é um *feedback* valioso, tanto para os professores como para os alunos, para que se possa melhorar o desempenho destes últimos (Sorour et al., 2015).

Uma grande parte das instituições académicas têm como meta de longo prazo aumentar o sucesso escolar dos alunos (Sorour et al., 2015). Tanto os professores como os investigadores defendem que a capacidade de prever o desempenho do aluno é extremamente importante em ambientes educacionais de modo a aumentar o sucesso no seu ambiente de aprendizagem (Reimers e Neovesky, 2015). Como tal, o investimento no de-

envolvimento de ferramentas que ajudem a rastrear e a avaliar o desempenho dos estudantes de uma dada instituição de ensino ao longo do tempo é de extrema importância.

1.2 MOTIVAÇÃO E OBJETIVOS

Nos últimos anos têm-se evidenciado algumas questões problemáticas no setor da educação, que são frequentemente realçadas pela considerável percentagem de insucesso e desistência académica. Um dos problemas apontados, frequentemente, pelos professores é que estes apenas têm acesso aos resultados finais alcançados pelos alunos, o que não revela todo o processo que estes seguiram. Desta maneira, os professores não conseguem atuar efetivamente como facilitadores para a aprendizagem da forma como pretendiam ou gostariam (Maldonado et al., 2012).

Tanto professores como pesquisadores e pais consideram extremamente importante ter uma visão global do progresso dos alunos e dos possíveis problemas que possam ocorrer, de modo a aumentar o desempenho e sucesso dos alunos através de um acompanhamento mais personalizado (Reimers e Neovesky, 2015). Desta forma, torna-se notória a importância do desenvolvimento de ferramentas que ajudem a rastrear e a avaliar o desempenho dos estudantes de uma dada instituição de ensino ao longo do tempo.

Uma das soluções propostas para atenuar este problema apresentado pelos professores recorre à análise da aprendizagem. A análise da aprendizagem consiste na avaliação das atividades dos alunos através da medição, coleção e análise dos dados relativos aos mesmos (Jiv, 2018). Esta avaliação visa melhorar e otimizar a aprendizagem e o ensino, através da previsão e monitorização do desempenho dos alunos (Reimers e Neovesky, 2015). Assim, poderá preencher a lacuna que existe entre as ciências da aprendizagem e a análise de dados (Jiv, 2018). Para tal, deverá ser considerada uma parte contínua e integrante do processo de aprendizagem, permitindo obter informações oportunas e específicas que podem auxiliar tanto os professores como os alunos. Deste modo, é esperado melhorar o processo de aprendizagem dos alunos como também o seu desempenho (Sorour et al., 2015).

O objetivo da aplicação de estratégias como essas é detetar, o mais cedo possível, eventuais questões problemáticas, como, por exemplo, sobrecarga de atividades, falta de material didático atualizado e de qualidade ou até falta de motivação dos alunos (Márquez-Vera et al., 2013b). Assim, torna-se possível auxiliar na tomada de decisões construtivas, informadas e conscientes, o que permite aos professores e alunos executar as tarefas de forma mais eficiente de modo a evitar ou reduzir o fracasso escolar (Avella et al., 2016).

Nesta dissertação, para implementar uma ferramenta de avaliação de desempenho dos estudantes, procurou-se analisar vários sistemas e modelos de avaliação existentes, com a motivação de estabelecer, inicialmente, um modelo de rastreio e de avaliação de desempenho. Depois, com a modelação e implementação do sistema de avaliação, com técnicas de mineração de dados - *Data Mining* (DM), procurou-se desenvolver um conjunto de *dashboards* adaptativos que permitissem a monitorização em tempo real e a análise do desempenho dos alunos com as melhores técnicas, modelos e componentes.

O sistema que foi desenvolvido permite a monitorização do desempenho dos alunos em avaliações de questionários – *quizzes*, realizados numa plataforma específica de *eLearning*, que integram um leque muito diver-

sificado de questões nos domínios e subdomínios envolvidos no processo de aprendizagem e avaliação dos alunos. Para além de uma análise individual, por aluno, é ainda possível fazer uma análise geral do desempenho dos alunos ao longo do tempo bem como do desempenho previsto, nos vários domínios e subdomínios. Futuramente, pretende-se que o sistema desenvolvido seja integrado na plataforma de *eLearning* Leonardo (Leonardo, 2018).

1.3 TRABALHO REALIZADO

Para fazer a implementação de uma ferramenta que ajude a rastrear e a avaliar o desempenho dos alunos de uma dada instituição de ensino ao longo do tempo é necessário ter em conta um conjunto de requisitos e funcionalidade concretos. Assim, nesta dissertação, primeiro procedeu-se à pesquisa de sistemas similares de avaliação do conhecimento de alunos já existentes, tendo sido selecionados seis sistemas distintos. De seguida, procedeu-se a uma análise comparativa dos sistemas selecionados, com o intuito de perceber quais seriam os pontos fortes e fracos dos *Learning Management System (LMS)* existentes e quais os problemas que ainda não foram solucionados. Numa fase seguinte, foi estudada a aplicação de vários modelos de *DM* na área da educação. A investigação dos modelos e técnicas de avaliação existentes, culminou na escolha de seis modelos distintos, que foram estudados e analisados detalhadamente em termos da sua aplicação e produção de resultados. Através de uma análise comparativa, foi possível definir qual o modelo de avaliação mais adequado e que foi desenvolvido nesta dissertação.

Depois da fase de estudo dos sistemas e de modelos de avaliação existentes iniciou-se o processo de desenvolvimento do sistema de avaliação de desempenho com a aplicação de técnicas de *DM*. Este desenvolvimento teve por base a metodologia *CRoss Industry Standard Process for Data Mining (CRISP-DM)*, desde a fase de conhecimento da área de aplicação até à modelação e análise dos resultados obtidos. Inicialmente, para o desenvolvimento do sistema foi realizada a angariação e o pré-processamento dos dados que tínhamos à nossa disposição, para de seguida ser feita a sua classificação através dos vários algoritmos selecionados. Depois procedeu-se à implementação da previsão do desempenho para 5, 10 e 15 dias. Por último, com o objetivo de analisar os dados, foram aplicadas técnicas de *clustering*. Após o desenvolvimento do sistema de avaliação procedeu-se à análise crítica dos resultados obtidos. Na última etapa, foram estudados os inúmeros componentes, modelos e técnicas de visualização de dados existentes com o intuito de desenvolver um conjunto de *dashboards* adequados para o problema em questão.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

Para além deste capítulo, esta dissertação inclui mais seis capítulos, que estão organizados da seguinte forma:

- Capítulo 2 - Sistemas de avaliação. Neste capítulo procedeu-se à pesquisa de sistemas similares de rastreio e avaliação de conhecimento de estudantes. Adicionalmente, apresenta-se as características e funcionalidades de seis sistemas bem como se faz a sua análise comparativa.

- Capítulo 3 - Análise e Avaliação de Desempenho. Aqui é analisada a aplicação de DM ao setor da educação e explicado os seus fundamentos. Para além disso, foi realizado um estudo dos modelos de avaliação de desempenho existentes, tendo sido explorados e descritos seis modelos distintos. Após essa análise, é feita uma breve comparação entre os vários modelos analisados.
- Capítulo 4 - Um Sistema para a Avaliação de Desempenho. Neste capítulo estão descritos todos os passos do desenvolvimento do modelo de avaliação de desempenho, nomeadamente: o processo de angariação e pré-processamento de dados, os modelos de classificação implementados, a previsão do desempenho e, por último, a implementação de *clustering*.
- Capítulo 5 - Análise dos Modelos Implementados. Após o desenvolvimento e implementação dos modelos, neste capítulo apresenta-se, em detalhe, os resultados obtidos em todos os passos do desenvolvimento e analisando-os com o intuito de definir o modelo de avaliação mais apropriado para o caso em estudo.
- Capítulo 6 - Visualização e Análise de Dados de Desempenho. A escolha dos componentes e modelos de visualização adequados para o objetivo é explicada detalhadamente neste capítulo. Complementarmente, apresentam-se os *dashboards* desenvolvidos e o seu modo de funcionamento com um conjunto de dados de uma plataforma de *eLearning*.
- Capítulo 7 - Conclusões e Trabalho Futuro. Neste capítulo é apresentado um resumo do trabalho realizado, bem como uma análise crítica do modelo de avaliação de desempenho desenvolvido e dos resultados obtidos. Termina-se este capítulo com uma exposição de algumas linhas de orientação para trabalho futuro.

SISTEMAS DE AVALIAÇÃO

2.1 UMA BREVE INTRODUÇÃO

Inúmeras instituições de ensino superior, tais como as universidades, procuram cada vez mais fornecer ferramentas de visualização de dados com utilidade, tanto aos seus professores como aos seus alunos. Contudo, de momento, ainda não existem soluções que atenuem todas as necessidades existentes no setor da educação. Uma solução existente e utilizada mundialmente são os sistemas de gestão de aprendizagem - LMS (Reimers e Neovesky, 2015). Existem diferentes expressões utilizadas para descrever estas aplicações informáticas educacionais, tais como sistemas de *e-Learning*, sistema de gestão de cursos - *Course Management System* (CMS) ou mesmo ambiente de aprendizagem virtual - *Virtual Learning Environment* (VLE) (Costa et al., 2012). Nestes sistemas, os alunos podem aceder aos conteúdos dos cursos em diferentes formatos, bem como interagir com professores e colegas, através de mensagens, fóruns, *chats*, videoconferência ou outros tipos de ferramentas de comunicação. Estas plataformas oferecem um conjunto de características configuráveis, de modo a permitir a criação de cursos, páginas de disciplinas, grupos de trabalho ou comunidades. Para além da dimensão pedagógica, possuem um conjunto de funcionalidades para atividades de registo, monitorização e avaliação de alunos e professores, o que permite fazer a gestão dos conteúdos de forma muito efetiva (Costa et al., 2012).

Nas últimas décadas, a tecnologia de gestão de aprendizagem revolucionou o ensino e a aprendizagem, uma vez que permitiu conceber e implementar sistemas como o Moodle (Moodle, 2020) ou a Blackboard (Blackboard, 2018), que permitem disponibilizar, gerir e monitorizar em tempo real todas as atividades dos alunos contribuindo para a melhoria da aprendizagem e do ensino (Park e Jo, 2015). Uma vez que os LMS não oferecem soluções reais de *dashboards*, as instituições de ensino superior começaram a investir no desenvolvimento deste tipo de “painéis” de visualização de dados, de forma que sejam úteis para os seus alunos e professores (Reimers e Neovesky, 2015).

2.2 SISTEMAS EXISTENTES

2.2.1 Moodle

O Moodle (Moodle, 2020) é um sistema de gestão de aprendizagem *open source*, lançado em 2001, que disponibiliza inúmeras ferramentas de comunicação, colaboração e avaliação. Atualmente, é uma das ferramen-

tas didáticas mais fáceis de utilizar e flexível disponível no mercado, uma vez que foi desenvolvido tendo em mente tanto a pedagogia como a tecnologia. Para além disso, possui uma excelente documentação e forte apoio à segurança e administração (Momani, 2012). O Moodle é utilizado em todo o mundo por universidades, escolas, empresas e professores independentes. Em 2016, era o sistema escolhido em 234 países com mais de 88,204,960 utilizadores registados (Keles e Özel, 2016).

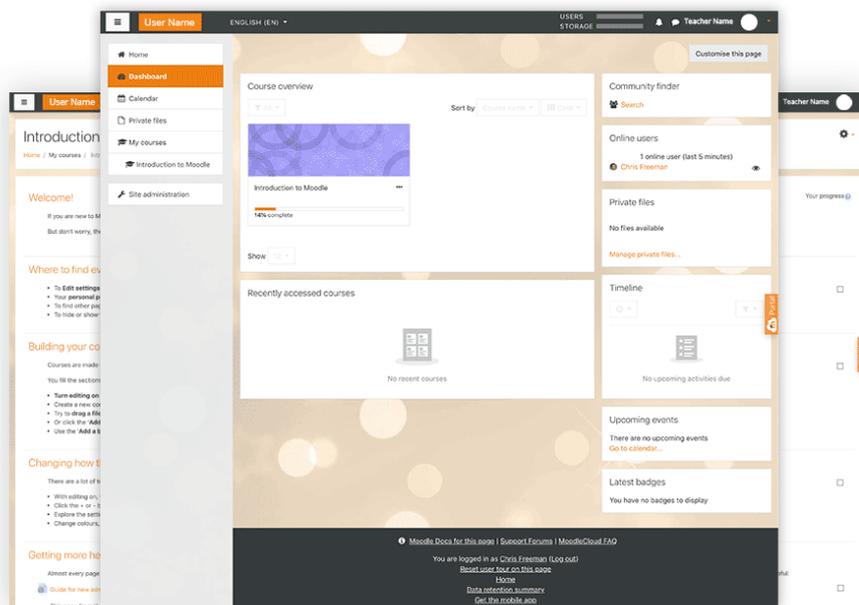


Figura 1: O ambiente de trabalho do Moodle

O Moodle é um sistema robusto, seguro e integrado, que disponibiliza um meio e ferramentas para criar ambientes de aprendizagem personalizados para os professores, administradores e alunos (Moodle, 2020). Como podemos observar na Figura 1, os professores podem criar e gerir cursos *online*, o que é fundamental para auxiliar também no ensino à distância, pois podem monitorizar os seus alunos, acedendo a estatísticas, bem como acompanhar todas as atividades realizadas (Momani, 2012).

O Moodle tem características que lhe permitem escalar para grandes implementações e centenas de milhares de estudantes, mas também pode ser utilizado por escolas primárias ou em ambientes de *hobbies* educativo. Muitas instituições utilizam-no como plataforma para realizar cursos totalmente *online*, enquanto algumas o utilizam simplesmente para aumentar os cursos presenciais (Reyes et al., 2009).

2.2.2 Blackboard

Em 2006, os proprietários das empresas Blackboard Learning System e WebCT decidiram fundir-se no Blackboard (Keles e Özel, 2016). O sistema possui uma grande variedade de produtos e serviços de software para programas educacionais. Além disso disponibiliza um ambiente de aprendizagem robusto, com grande ampli-

tude de funcionalidades, uma interface moderna, intuitiva e totalmente reativa. Estas características permitem ao Blackboard proporcionar uma experiência de ensino e aprendizagem mais simples e poderosa, que vai além dos sistemas tradicionais de gestão da aprendizagem, como podemos analisar na Figura 2 (Blackboard, 2018).

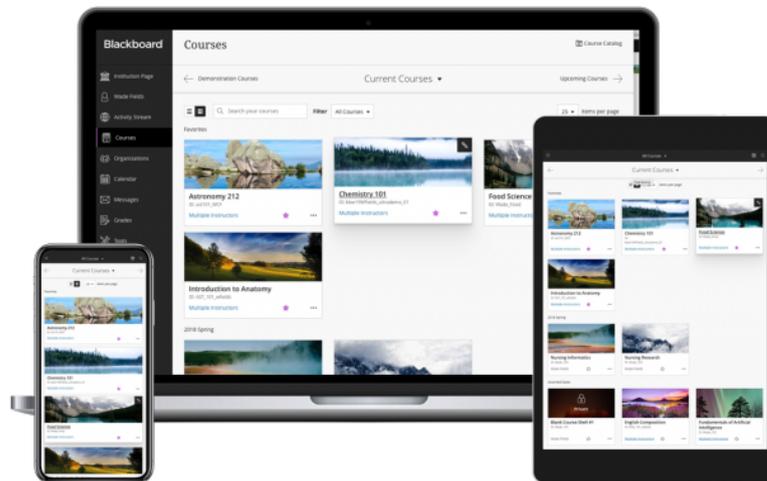


Figura 2: O ambiente de trabalho do Blackboard

No Blackboard, os professores e os alunos conseguem, de uma forma intuitiva, interagir com cursos, conteúdos, instrutores e os seus colegas (Blackboard, 2018). O sistema oferece *feedback* útil para os professores recorrendo a dashboards que contém estatísticas instantâneas sobre o acesso dos alunos, notas e participação (Momani, 2012). As características de disponibilidade, *feedback* rápido, comunicação melhorada e a oportunidade de interagir com os instrutores, destacaram o Blackboard das restantes aplicações LMS. Isso fez com que fosse considerado como um dos mais populares e comercializáveis sistemas de aprendizagem colaborativa adotados no ensino superior (Almekhlafy, 2020).

2.2.3 Claroline

O Claroline (Claroline, 2020) é uma plataforma colaborativa de aprendizagem *open source*, que foi desenvolvida tendo em conta a experiência e necessidades pedagógicas dos professores. Esta ferramenta permite que, centenas de organizações em todo o mundo, desde universidades, escolas e empresas, possam criar e gerir cursos. A plataforma, lançada em 2000 pela UCL - Universidade Católica de Lovaina, é agora usada em mais de 80 países e está disponível em mais de 30 idiomas (Reyes et al., 2009). Para uma melhor coordenação do desenvolvimento da plataforma e para promover a sua utilização numa comunidade de utilizadores, em 2007 foi criado o Consórcio Claroline, uma associação internacional sem fins lucrativos (Reyes et al., 2009).

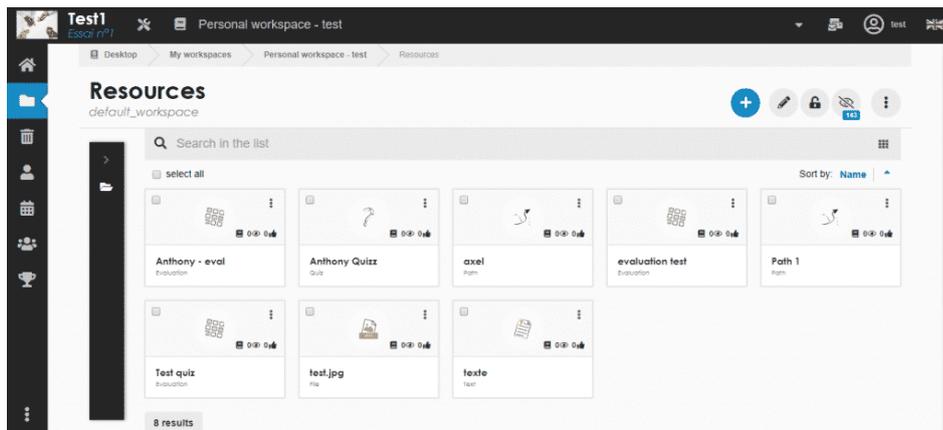


Figura 3: O ambiente de trabalho do Claroline

O foco do Claroline é a aprendizagem, pelo que tudo neste sistema é pensado para facilitar a integração do conhecimento. Assim, o número de recursos disponibilizados é considerável, o que o torna também adaptável às necessidades das organizações. Este sistema disponibiliza *chat* e fórum de discussão, para além de permitir criar e gerir a aprendizagem através de *dashboards* (Claroline, 2020). Na Figura 3, podemos analisar como é o *workspace* deste sistema e alguns dos recursos que tem disponíveis.

2.2.4 ATutor

O ATutor (ATutor, 2020) é um sistema de gestão de aprendizagem *open source* escolhido por muitas universidades, centros de pesquisa e instituições de ensino (Keles e Özel, 2016). O sistema foi projetado para desenvolver cursos *online* para além de criar e distribuir conteúdo *e-learning*, bem como disponibiliza recursos tanto para estudantes como professores, entre os quais a monitorização das atividades e recursos (ATutor, 2020).

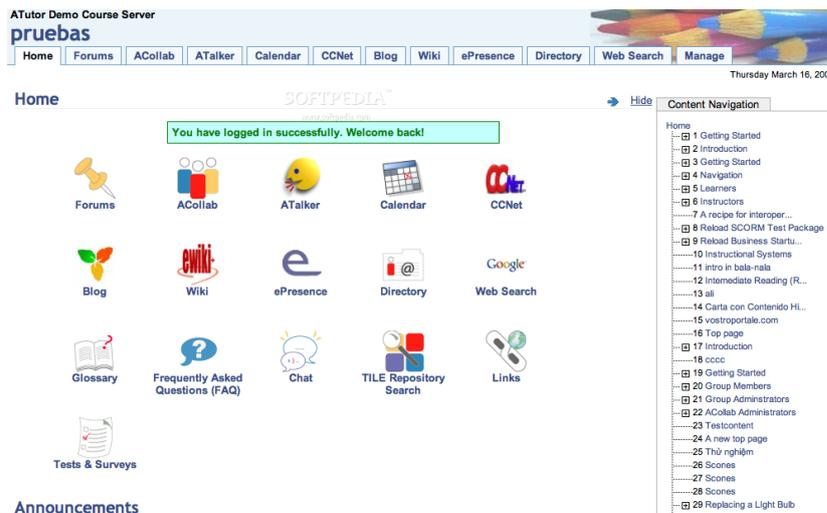


Figura 4: O ambiente de trabalho do ATutor

O ATutor possui uma ferramenta que permite a troca de mensagens instantâneas num *chat* em tempo real e também a troca de mensagens privadas, bem como um *whiteboard*. É considerado um sistema único, porque oferece espaço para os utilizadores desenvolverem uma rede de contactos, criarem um perfil de rede, juntarem-se a grupos de interesse, partilharem imagens e discutirem sobre os materiais pedagógicos disponíveis (Cavus e Zabadi, 2014). Na Figura 4, podemos verificar alguns dos recursos que o ATutor coloca à nossa disposição. Para além disso, o sistema é destacado pela sua acessibilidade aos recursos fornecidos, que são adaptados para pessoas com deficiência visual. A adequação ao ensino, uma vez que é adaptável aos diversos cenários de ensino, e a adaptação a uma grande variedade de tecnologias, com suporte *web* e *mobile*, são características a realçar neste sistema (Reyes et al., 2009). Porém, o ATutor, apesar de apresentar uma grande usabilidade, carece da falta de várias outras características que ainda se encontram em desenvolvimento (Cavus e Zabadi, 2014).

2.2.5 Dokeos

A Dokeos é uma das maiores e mais reconhecidas empresas dedicadas a *LMS open source*. O seu sistema de aprendizagem é reconhecido por empresas multinacionais e universidades em cerca de 60 países com suporte para mais de 30 línguas, num total de 1,297,000 utilizadores (Reyes et al., 2009).

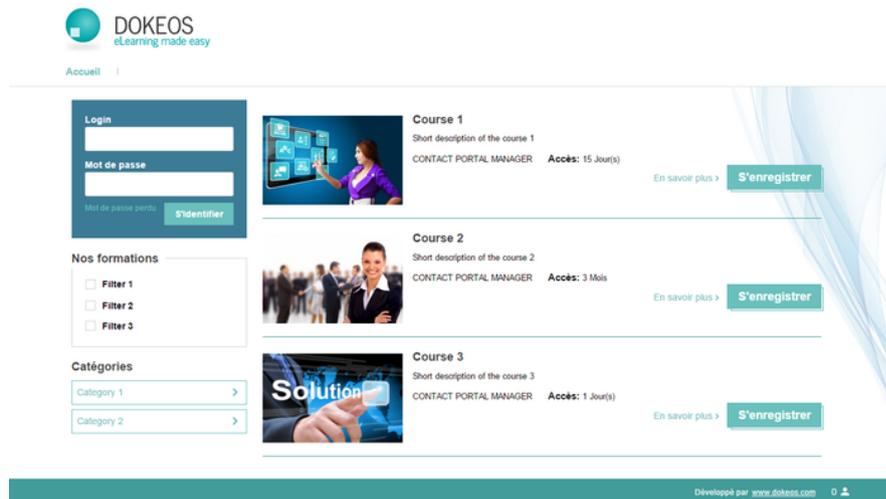


Figura 5: O ambiente de trabalho do Dokeos

Esta plataforma de *eLearning* disponibiliza ferramentas de videoconferência e um fórum de discussão para que os estudantes possam interagir entre si. Contudo não tem um *whiteboard* nem *chat* em tempo real. Na Figura 5 podemos ver algumas das ferramentas fornecidas pelo sistema para gerir utilizadores, cursos e sessões (Cavus e Zabadi, 2014). Além disso, o sistema *e-learning* permite criar facilmente programas poderosos e eficazes que treinam, validam e comunicam. É possível acompanhar e monitorizar o progresso dos participantes através da análise dos resultados de avaliações (Docebo, 2020).

2.2.6 Docebo

Outra plataforma *e-learning LMS open source* é Docebo (Docebo, 2020). Esta sistema é usado tanto no setor da educação como no setor corporativo (Reyes et al., 2009). Suporta diferentes modelos didáticos, em 18 idiomas em mais 10 países, contando com mais de 53 funcionalidades diferentes (Keles e Özel, 2016).

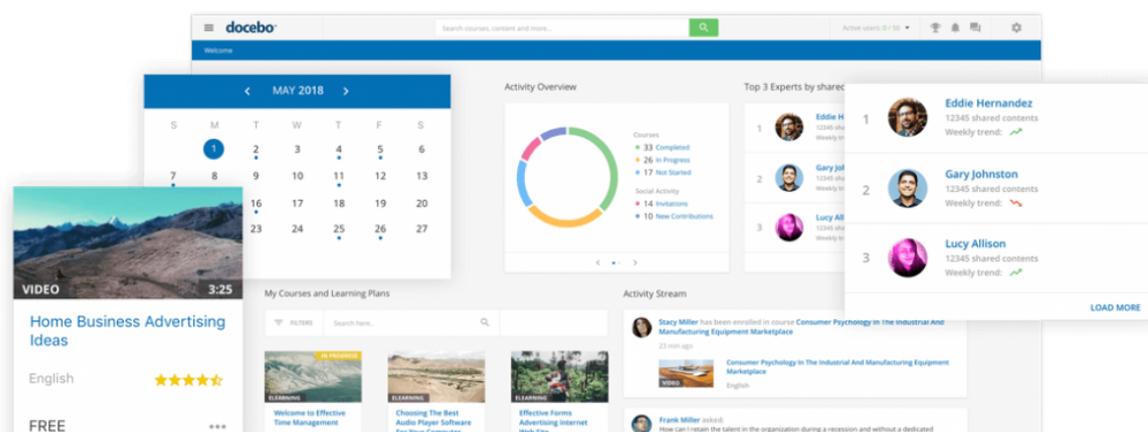


Figura 6: O ambiente de trabalho do Docebo

Na Figura 6 podemos verificar que para além de permitir a criação e gestão de cursos ainda disponibiliza um fórum de discussão, *wiki*, *chat* e videoconferência (Reyes et al., 2009). O Docebo foi projetado para habilitar, aprimorar e transformar a aprendizagem online para as empresas modernas, proporcionando um *design* preciso, funcional e flexibilidade em escala (Docebo, 2020).

2.3 ANÁLISE COMPARATIVA

Todos os LMS estudados permitem a criação e gestão de cursos e *quizzes*, disponibilizando a análise e monitorização das atividades dos estudantes. Apesar de apresentarem uma solução para algumas das lacunas existentes no sistema de ensino e apresentarem suporte tanto para plataformas *web* como para plataformas móveis, todos eles apresentam, naturalmente, características e recursos diferentes, que os distinguem. Dos sistemas apresentados apenas o Blackboard e o Docebo não são *open source*, o que significa que para serem utilizados precisamos de ter uma licença comercial. Alguns dos recursos, como o fórum de discussão e o *chat* são comuns a todos os sistemas, mas a funcionalidade de vídeo conferência não está disponível no Claroline. Em termos de *whiteboards*, apenas o Docebo não tem este serviço.

A comparação dos seis LMS mostra que o Moodle e o ATutor possuem as melhores ferramentas de comunicação com interface amigável, o que permite um acesso fácil e eficaz à informação (Cavus e Zabadi, 2014). Contudo, o Blackboard e o Moodle são bastante semelhantes em termos de recursos administrativos, colaboração e métodos de instrução. Em termos de interface, o Docebo aparenta ter melhores funcionalidades e é menos complexo do que o Moodle. O Docebo destaca-se pelos seus recursos administrativos e desenvolvimento de cursos (Keles e Özel, 2016). O sistema de gestão de aprendizagem mais complexo é o Claroline, pelo que a sua utilização é mais difícil (Cavus e Zabadi, 2014).

A adaptabilidade não é alcançada na totalidade em nenhum dos sistemas comparados, tal como é possível verificar na Tabela 1. Nesta tabela podemos ver um resumo de algumas das características e funcionalidades

disponibilizadas por cada um dos sistemas que vimos. Contudo, todos têm fóruns de discussão que permitem desenvolver versões melhoradas, o que é um dos aspectos positivos dos LMS (Cavus e Zabadi, 2014).

| | Moodle | Blackboard | Claroline | ATutor | Dokeos | Docebo |
|----------------------|---------------|-------------------|------------------|---------------|---------------|---------------|
| Pago | Não | Sim | Não | Não | Não | Sim |
| Monitorização | Sim | Sim | Sim | Não | Sim | Sim |
| Quizzes | Sim | Sim | Sim | Sim | Sim | Sim |
| Chat | Sim | Sim | Sim | Sim | Sim | Sim |
| Conferências | Sim | Sim | Não | Sim | Sim | Sim |
| WhiteBoard | Sim | Sim | Sim | Sim | Não | Sim |
| Open Source | Sim | Não | Sim | Sim | Sim | Não |
| Suporte web e mobile | Sim | Sim | Sim | Sim | Sim | Sim |

Tabela 1: Análise comparativa de LMS

ANÁLISE E AVALIAÇÃO DO DESEMPENHO DOS ESTUDANTES

3.1 CONTEXTUALIZAÇÃO E OBJETIVOS

A avaliação de desempenho dos alunos é apontada por especialistas como sendo uma parte contínua e integrante do processo de aprendizagem dos alunos. A monitorização do desempenho envolve avaliações que têm um papel vital no fornecimento de informação orientada para ajudar estudantes, professores, administradores, e decisores políticos a tomar decisões (Ogor, 2007).

A grande quantidade de dados atualmente existentes nas bases de dados dos estudantes excede a capacidade humana de analisar e extrair a informação mais útil sem a ajuda de técnicas de análise automatizada (Algarni, 2016). Estes fatores, em mudança na educação contemporânea, levaram então à procura de monitorização eficaz e eficiente do desempenho dos estudantes nas instituições de ensino. As técnicas tradicionais de medição e avaliação começaram a ser substituídas por técnicas de DM (Ogor, 2007). Os processos de DM ou de rastreamento detalhado das atividades realizadas pelos estudantes ao longo do tempo é uma das formas que existe para sustentar a análise da aprendizagem dos estudantes (Teasley, 2017). A metodologia de DM permite que, ao mesmo tempo que se extrai padrões úteis e válidos do ambiente da base de dados do ensino superior, se contribua para assegurar proactivamente que os estudantes maximizem o seu desempenho académico (Ogor, 2007).

Os processos numa intervenção analítica de aprendizagem típica são bastante mais simples e expeditos, se para além da aplicação de DM utilizarmos ferramentas de visualização de dados adequadas, tais como os *dashboards* (painéis interativos para a visualização de dados) (Jiv, 2018). Estes painéis são considerados mais do que uma simples coleção de indicadores e de gráficos, com as formas e interfaces mais variados, capazes de fornecerem visualmente os elementos de dados necessários para responder às mais diversas perguntas, de acordo com as dimensões de análise que neles possam estar refletidas (Bolos et al., 2014b).

Da perspetiva de quem toma decisões, os *dashboards* fornecem uma forma bastante fácil, rápida e útil de visualizar informação importante, consolidada e organizada, segundo uma ou mais vertentes de análise, envolvendo uma ou mais medidas de análise (Teasley, 2017). Através do aproveitamento dos recursos visuais é possível identificar visualmente tendências, padrões e anomalias e, desta forma, orientar na tomada de decisões (Few, 2006a).

Desde que se possua a informação necessária, estes painéis de visualização de dados permitem, em tempo real, identificar facilmente os problemas que possam estar a ocorrer no processo de aprendizagem ou analisar

os resultados da aplicação deste ou daquele método de ensino (Reimers e Neovesky, 2015). Esta monitorização do progresso da aprendizagem dos alunos ajuda os professores a desempenharem as suas funções de maneira eficaz, tanto na facilitação de aprendizagem, como no fornecimento de *feedbacks*, avaliação e classificação (Park e Jo, 2015).

As universidades podem recorrer aos *dashboards* para gerir o desempenho dos alunos, definindo métricas e indicadores de desempenho que são monitorizados ao longo do tempo. Deste modo, nos próximos anos prevê-se que estes painéis de visualização de dados se tornem essenciais nas universidades (Muntean et al., 2010). A monitorização poderá ser interativa, histórica, personalizada e analítica, o que reflete os padrões de aprendizagem, desempenho e interações dos alunos (Park e Jo, 2015).

3.2 MODELOS DE AVALIAÇÃO DO DESEMPENHO DE ESTUDANTES

A aplicação de modelos de DM permite identificar o mais cedo possível quais os alunos que precisam de um acompanhamento especial e, desta forma, investir em estratégias que visem melhorar o processo de ensino (Kumar e Pal, 2011). Esta solução é considerada muito promissora para reduzir o fracasso e abandono escolar, através da sua previsão antecipadamente (Márquez-Vera et al., 2013a).

O DM é uma ferramenta poderosa de inteligência artificial - *Artificial Intelligence (AI)*, que pode descobrir informação útil através da análise dos dados de muitos ângulos ou dimensões, categorizar essa informação, e resumir as relações identificadas na base de dados. Subsequentemente, esta informação ajuda a tomar ou a melhorar as decisões. Nas soluções de DM, os algoritmos podem ser utilizados independentemente ou em conjunto para alcançar os resultados desejados.

No decorrer dos últimos anos, o uso crescente da tecnologia nos sistemas educativos tornou disponível uma grande quantidade de dados, o que contribuiu para um aumento no interesse da aplicação de técnicas de mineração de dados educacionais - *Education Data Mining (EDM)* para analisar e prever o desempenho dos alunos (Sorour et al., 2015).

O processo de aplicação de DM para sistemas de ensino pode ser interpretado segundo diferentes pontos de vista. De um ponto de vista educacional e experimental, pode ser visto como um ciclo iterativo de formação de hipóteses, testes e aperfeiçoamento, tal como podemos verificar na Figura 7 (Romero e Ventura, 2013).

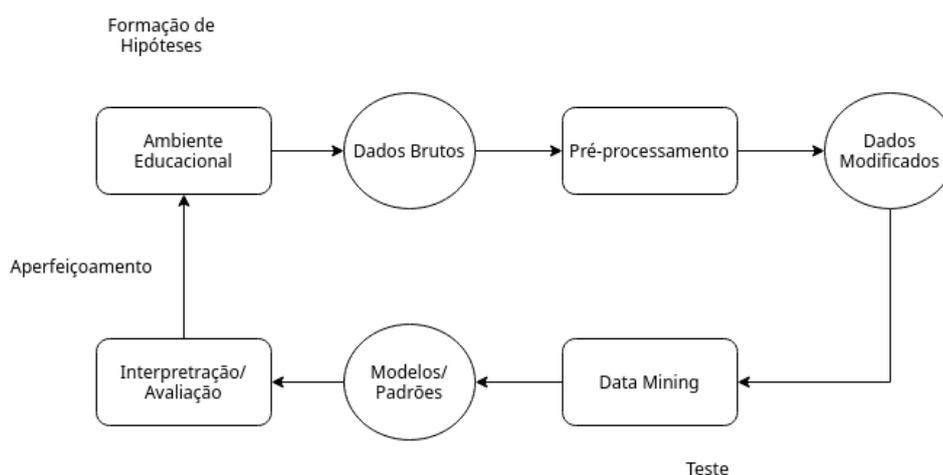


Figura 7: Descoberta de conhecimento educacional e processo de DM – figura adaptada de (Romero e Ventura, 2013)

Neste processo, o objetivo não é apenas transformar dados em conhecimento, mas também filtrar o conhecimento extraído para a tomada de decisões sobre como modificar o ambiente educacional, de modo a melhorar a aprendizagem do aluno. Este é um tipo de avaliação formativa de um programa educativo, enquanto ainda está em desenvolvimento, e com a finalidade de melhorar continuamente o programa.

A EDM é uma disciplina emergente que visa desenvolver métodos para explorar os tipos de dados exclusivamente provenientes do contexto educacional e utilizar estes métodos para compreender melhor os estudantes e os ambientes em que eles aprendem (Romero et al., 2008b). Estes diferem dos típicos métodos de DM uma vez que quando utilizada explicitamente, é responsável pela hierarquia multinível e carece de dados indutivos independentes (Algarni, 2016). Esta pode ser esboçada como a combinação de três áreas principais, (Figura 8): ciências da computação, educação e estatística. A interseção dessas três áreas também torna outras subáreas intimamente relacionadas com a EDM, como: a educação baseada em computador, DM, *Machine Learning* (ML) e análise de aprendizagem - *Learning Analytics* (LA).

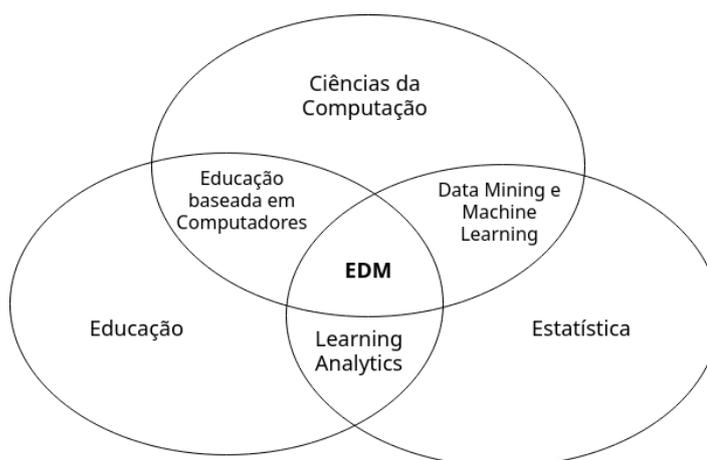


Figura 8: Principais áreas relacionadas com a EDM – figura adaptada de (Romero e Ventura, 2013)

Recorrendo a estas técnicas é possível ter acesso a uma quantidade significativa de informação relevante uma vez que fornece uma imagem mais clara dos alunos e dos seus processos de aprendizagem. À semelhança de outros processos de extração de técnicas de DM, a EDM extrai informação interessante, interpretável, útil e inovadora a partir de dados educativos.

A sua utilidade é notória em diversas áreas, incluindo a identificação de alunos em risco, identificação de necessidades de aprendizagem prioritárias para diferentes grupos de alunos, aumento das taxas de graduação, avaliação eficaz do desempenho institucional, maximização dos recursos do campus e otimização da renovação curricular da disciplina (Algarni, 2016).

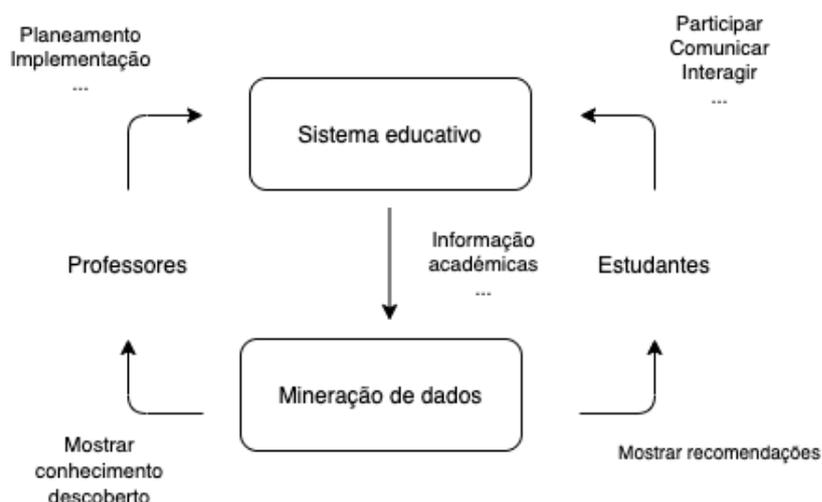


Figura 9: Mineração de dados educacionais – figura adaptada de (Romero e Ventura, 2007)

A Figura 9 apresenta o cenário básico da utilização da EDM, no qual existem grupos de interessados, tais como estudantes ou professores, que podem beneficiar destas técnicas de diferentes formas. No caso dos

estudantes, estes podem receber conselhos ou recomendações de recursos ou tarefas que são adequadas para complementarem o conhecimento atual e atingirem os objetivos da aprendizagem. Enquanto os professores podem analisar a eficácia da metodologia aplicada e dos recursos fornecidos através da observação dos dados analisados (Romero e Ventura, 2007).

3.3 ANÁLISE DE MODELOS DE AVALIAÇÃO

O primeiro modelo estudado trata-se de uma nova abordagem baseada em técnicas de mineração de texto para prever o desempenho dos alunos recorrendo aos métodos de análise semântica latente - *Latent Semantic Analysis* (LSA) e *k-means clustering*. Estes métodos permitem a extração de informação semântica dos comentários escritos livremente pelos alunos. Para estipular situações de aprendizagem os comentários são categorizados aplicando o método PCN, no qual a aprendizagem antes das aulas é representada por P, a compreensão e aproveitamento durante as aulas C e o plano de atividades de aprendizagem posteriormente às aulas N. Este modelo é constituído por cinco fases: recolha dos comentários, preparação de dados com LSA, fase de treino aplicando o método *k-means clustering*, fase de teste e deteção de ruídos (Sorour et al., 2015).

A recolha dos comentários é executada com a aplicação do método PCN. De seguida, são aplicados 3 diferentes passos na preparação dos dados. Primeiramente é feita uma análise morfológica, de seguida calculada a frequência das palavras e, por fim, a aplicação do método LSA para analisar padrões e relações. Na fase de treino é aplicado o algoritmo *k-means clustering* com 5 clusters e a fase de testes foi realizada recorrendo a validação cruzada – *cross validation* (CV) com 10-fold com 90% dos comentários para treino e 10% de teste. Por fim, é verificada a existência de outliers (Sorour et al., 2015).



Figura 10: Processo de mineração de dados

Nos vários modelos estudados, o processo de mineração de dados (Figura 10) permite identificar, extrair e avaliar as variáveis no processo de aprendizagem. Vários algoritmos e técnicas, como classificação, regressão, redes neuronais, algoritmos genéticos, entre outros, são aplicados para extrair conhecimento de bases de dados (Kumar e Pal, 2011).

No segundo modelo estudado de modo a prever o desempenho no final do semestre, o processo de EDM aplicado foi constituído por seis etapas distintas, preparação de dados, seleção e transformação, classificação, medição da impureza e aplicação do critério de separação. Neste estudo foram aplicadas as árvores de decisão - *Decision Tree* (DT), que são normalmente utilizadas para obter informação para efeitos de tomada de decisões. Existem três tipos de algoritmos de aprendizagem amplamente utilizados, ID3, ASSISTANT e C4.5. Neste caso

foi optado por aplicar o algoritmo simples de aprendizagem em DT, o ID3. Neste modelo a previsão é realizada através da análise de várias variáveis da avaliação e resultados obtidos pelos alunos, tais como os resultados dos exames e dos laboratórios (Kumar e Pal, 2011).

O terceiro modelo analisado, também de EDM, considera os fatores que afetam o desempenho dos alunos e é composto pelas mesmas quatro fases apresentadas anteriormente na Figura 10: seleção de todos os dados sobre os alunos, pré-processamento, aplicação de métodos de classificação e, por fim, a interpretação dos resultados obtidos. Neste modelo são aplicados diferentes algoritmos de DM que são executados com variáveis em estudo diversas, desde interação social, atividades extracurriculares ou bases acadêmicas. Os algoritmos aplicados são DT, redes neurais - *Neural Network* (NN), *Naive Bayes* e *K-Nearest Neighbor* para no final os resultados serem comparados e, desta forma, ser selecionado o algoritmo que produz os melhores resultados. As técnicas de classificação foram selecionadas tendo por base a reputação na literatura de mineração de dados recentemente publicada e na superioridade de previsão em problemas deste tipo. As DT apresentam como vantagem a detecção e interpretação das regras criadas sem esforço devido ao seu formato em forma de árvore, o que diminui a ocorrência de erros. Neste estudo, foi utilizado o algoritmo C4.5, um algoritmo de alta classificação na pesquisa de prospecção de dados. No caso das NN estas têm a capacidade excepcional de desenvolver significado a partir de dados complexos. A percepção MLP foi o modelo selecionado uma vez que é o mais famoso para objetivos de classificação acadêmica. O último modelo selecionado foi o *Naive Bayes* que é uma das abordagens mais simples para a construção de um modelo de classificação (Shahiri et al., 2015).

O objetivo do quarto modelo foi classificar os alunos com notas finais iguais em grupos diferentes dependendo das atividades realizadas na plataforma Moodle (Moodle, 2020). Para avaliar o desempenho e a utilidade dos diferentes algoritmos de classificação, o modelo tem por base os dados do sistema *e-learning*. Este modelo também é constituído pelas fases típicas do processo de DM apresentadas na Figura 10. Os dados analisados têm por base a interação e uso dos alunos, posteriormente na fase de pré-processamento é realizada a discretização dos valores numéricos e o balanceamento dos dados. Os algoritmos aplicados incluíram DT, alguns dos mais conhecidos algoritmos são o C4.5 e CART, indução de regras que é uma área de aprendizagem de máquinas em que as regras de produção *if-then* são extraídas de um conjunto de observações, as NN que também podem ser utilizadas para indução de regras, entre outros. No problema educacional em questão era muito importante que o modelo de classificação obtido fosse de fácil utilização, para que os professores pudessem tomar decisões, pelo que os algoritmos foram selecionados tendo em conta se iriam produzir modelos compreensíveis e de fácil interpretação. Estas técnicas foram aplicadas separadamente, mas também podem ser aplicadas em conjunto a fim de obter informações interessantes de forma mais eficiente e rápida (Romero et al., 2008a).

No quinto modelo em estudo foi implementado o processo de EDM numa plataforma de *e-learning*. Este é constituído pelas quatro etapas do processo de DM, que pode ser consultado na Figura 10. Numa primeira etapa é feita a seleção dos dados, neste caso em específico da plataforma Moodle. De seguida é realizado o pré-processamento dos dados que engloba a seleção dos dados bem como a sua discretização e transformação para o tipo de dados pretendido. Depois são aplicados os algoritmos de DM para construir e executar o modelo. Os algoritmos aplicados são agrupados em diferentes categorias, estatísticas, visualização, *clustering*, classificação e regra de associação.

A visualização pretende ajudar os utilizadores a compreender e analisar grandes quantidades de dados. Por outro lado, as estatísticas de utilização dos alunos são frequentemente o ponto de partida para a avaliação num sistema de *e-learning*. O *clustering* é um processo de agrupamento de objetos em classes de objetos semelhantes, o algoritmo aplicado foi o *k-means* que é considerado um dos algoritmos de agrupamento mais simples e mais populares. Existem vários classificadores disponíveis, mas neste estudo, foi aplicado a árvore de decisão com o algoritmo C4.5 para caracterizar os alunos que passaram ou reprovaram. As regras de associação permitem descobrir relações entre atributos em bases de dados, produzindo declarações *if-then* relativamente aos valores de atributo. Estas são as mais gerais e conhecidas técnicas de DM, contudo existem outras técnicas de DM que também são utilizadas nos sistemas *e-learning*, tais como mineração de padrões sequenciais, mineração de textos, análise de *outliers* e análise de redes sociais. Por fim, é possível interpretar e avaliar os resultados (Romero et al., 2008b).

O último modelo de EDM pretendeu, através da aplicação de técnicas de EDM, prever o fracasso e abandono escolar. Para tal, procura identificar quais os fatores, tanto pessoais como académicos, que influenciam o fracasso dos alunos. Neste modelo a metodologia é semelhante aos apresentados anteriormente, porém no pré-processamento dos dados, além da limpeza dos dados - *data cleaning*, a transformação de variáveis e partição dos dados - *data partitioning*, foram aplicadas técnicas como a seleção de atributos e balanceamento dos dados. Neste modelo é proposta a aplicação de diferentes técnicas de DM uma vez que este é um problema complexo e de grandes dimensões. No processo de DM foram utilizados cinco algoritmos de indução de regras (JRip, NNge, OneR, Prism e Ridor) e cinco de DT (J48, C4.5, SimpleCart, ADTree e RandomTree). As DM e algoritmos de indução de regras são bastante utilizadas por serem técnicas de classificação que fornecem uma explicação para o resultado da classificação e pode ser utilizado diretamente para a tomada de decisões. Para além disso, uma das vantagens das DT é que podem ser diretamente transformadas num conjunto de regras *if-then*, que é uma das formas mais populares de representação de conhecimento devido à sua simplicidade e compreensibilidade. Desta forma, um utilizador mesmo que não seja experiente na área de DM, pode utilizar diretamente a saída obtida por estes algoritmos para detetar estudantes com problemas (Márquez-Vera et al., 2013a).

3.4 ANÁLISE COMPARATIVA

Todos os seis modelos estudados tinham como objetivo prever o desempenho dos alunos. Para tal, foram aplicadas técnicas de DM, como foi o caso do modelo para classificação dos alunos, e métodos de EDM nos restantes. Estes diferem na medida em que analisam diferentes fatores que podem afetar o desempenho e também são aplicadas diferentes metodologias, desde as etapas do pré-processamento, bem como os algoritmos testados, tal como podemos verificar na Tabela 2. O primeiro modelo em estudo, de análise de aprendizagem, aparenta ser o menos apropriado para o caso em estudo uma vez que apenas tem em consideração comentários. Os modelos de previsão do desempenho e de previsão do fracasso e abandono escolar, apresentam uma grande variedade de variáveis em análise, englobando fatores extracurriculares e sociais, enquanto os restantes têm um maior foco no percurso escolar dos alunos, apresentando assim mais semelhanças com o problema a

solucionar. Os modelos tiveram por base as etapas constituintes do processo de DM (Figura 10), sendo depois adaptados. A maior parte dos modelos estudados tinha em comum a aplicação de diferentes algoritmos de DT e de indução de regras, contudo em todos foram aplicados diversos algoritmos de classificação diferentes com o intuito de analisar os resultados obtidos e selecionar o melhor selecionador.

| | Análise aprendi- zagem | Análise desem- penho | Previsão desempenho | Classificação de alunos | Sistemas de aprendi- zagem | Previsão fracasso/ abandono escolar |
|------------|---------------------------------------|-------------------------------------|--------------------------------|--|---|--|
| Previsão | Sim | Sim | Sim | Sim | Sim | Sim |
| EDM | Sim | Sim | Sim | Não | Sim | Sim |
| Variáveis | 1 | 8 | 36 | 11 | 13 | 77 |
| Etapas | 5 | 7 | 3 | 3 | 4 | 4 |
| Algoritmos | 1 | 1 | 3 | 5 | 9 | 10 |

Tabela 2: Análise comparativa dos modelos estudados

UM SISTEMA PARA A AVALIAÇÃO DE DESEMPENHO

4.1 CARACTERIZAÇÃO DO MODELO DE AVALIAÇÃO

O **DM** é um processo criativo que requer uma série de diferentes competências e conhecimento. Contudo, atualmente, não existe uma *framework standard* para a realização destes projetos. É necessário o desenvolvimento de uma abordagem padrão que ajude a traduzir problemas empresariais em tarefas de mineração de dados, ou seja, que sugira transformações de dados apropriadas e técnicas.

O *CRoss Industry Standard Process for Data Mining (CRISP-DM)* (Wirth, 2000) aborda parte destes problemas uma vez que este define um modelo do processo de modo a fornecer uma *framework* para a realização de projetos de **DM** que seja independente tanto do sector industrial como da tecnologia utilizada.

O modelo genérico do processo é útil para o planeamento, comunicação dentro e fora da equipa do projeto e documentação. Para além disso, fornece uma excelente base para o desenvolvimento de um modelo de processo especializado constituído por vários passos descritos ao pormenor e indica conselhos práticos para todas estas etapas. Desta forma, o processo torna os grandes projetos de **DM** menos dispendiosos, mais fiáveis, repetíveis, fáceis de gerir e mais rápidos.

O ciclo de vida da **DM** é decomposto em seis fases: compreensão do negócio, entendimento de dados, preparação de dados, modelação, avaliação e implementação. Na Figura 11 é apresentada a sequência das várias fases, que não é rigorosa. As setas neste caso indicam as dependências mais importantes e frequentes entre fases e o círculo exterior simboliza a natureza cíclica do processo de **DM**. A extração de dados não termina assim que uma solução é encontrada e implementada uma vez que são aprendidas lições durante o processo de implementação da solução que podem desencadear novas questões empresariais.

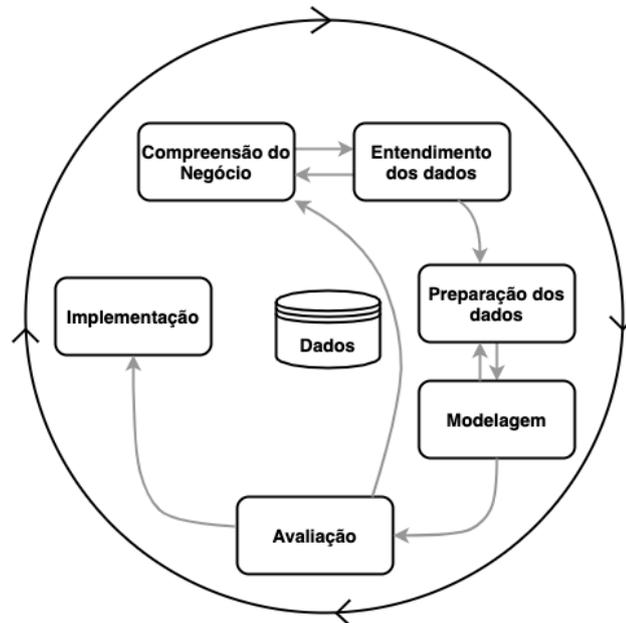


Figura 11: Modelo CRISP-DM – figura adaptada de (Wirth, 2000)

A primeira fase da metodologia CRISP-DM é a compreensão do negócio - *business understanding* - que se centra na compreensão dos objetivos e requisitos do projeto, seguida da conversão deste conhecimento numa definição detalhada do problema de DM bem como num plano preliminar para alcançar os objetivos.

De seguida, a fase de compreensão de dados - *data understanding* começa com a recolha inicial de dados e prossegue com a análise dos mesmos para que haja uma familiarização com os dados e identificar eventuais problemas de qualidade.

A fase de preparação de dados - *data preparation* - abrange todas as etapas de limpeza e transformação para construir o conjunto de dados final a partir dos dados em bruto iniciais, desde a limpeza dos dados - *data cleaning* até à redução dos mesmos - *data reduction*.

Em seguida, na fase de modelação - *modeling* - são seleccionadas e aplicadas diversas técnicas e analisados os parâmetros de modo a obter os valores ótimos. Esta etapa pode ser executada várias vezes, sendo até possível voltar a etapas anteriores para ajustar os dados de acordo com os requisitos da técnica a ser utilizada. Existem diversos algoritmos sendo que para este caso foram seleccionados três diferentes para serem treinados, validados e testados. Os modelos seleccionados foram as DT, *Random Forest* (RF) e *Support Vector Machines* (SVM). Esta escolha teve por base a sua reputação na literatura recentemente publicada (Capítulo 3) sobre DM e a sua superioridade em problemas deste tipo de previsão.

Na avaliação do projeto - *evaluation* - estão já construídos os modelos com alta qualidade, da perspetiva da análise dos dados. No final desta fase, deve ser tomada uma decisão sobre a utilização dos resultados da prospeção dos dados sendo feita uma análise dos modelos através da *accuracy* e da matriz de correlação.

Na fase final, que é a implementação – *deployment* - os conhecimentos adquiridos terão de ser organizados e apresentados de forma que o cliente possa utilizá-los (Wirth, 2000).

Esta foi a metodologia aplicada no desenvolvimento desta dissertação ao longo de todas as fases, desde o conhecimento da área até à modelação e análise dos resultados.

4.2 ANGARIAÇÃO E PRÉ-PROCESSAMENTO DE DADOS

Os dados colecionados são compostos pela avaliação obtida por vários estudantes num determinado período de tempo em *quizzes*, que abrangem todas as matérias selecionadas e avaliadas estando as questões agrupadas por domínios e subdomínios. Os alunos ao longo do ano letivo estudam de forma regular e são realizadas avaliações de forma contínua, pelo que a análise das questões e respostas permite analisar e avaliar o desempenho dos estudantes. No problema em questão foram analisados os dados por aluno, domínio e subdomínio das questões.

Para a aplicação de técnicas de classificação e previsão é necessário o pré-processamento de dados. Nesta fase é feita uma análise cuidadosa das variáveis e dos valores correspondentes para eliminar quaisquer anomalias e garantir a qualidade e confiabilidade das informações disponíveis. O pré-processamento é constituído por 4 etapas distintas: limpeza de dados (*data cleaning*), integração dos dados (*data integration*), transformação (*data transformation*) e redução de dados (*data reduction*).

O processo de limpeza de dados é caracterizado pelo processamento de valores em falta - *missing values*, dados ruidosos - *noisy data* e dados inconsistentes - *inconsistent data*. Os dados reais podem ser incompletos, com a existência de dados ausentes, ruidosos, se existirem erros aleatórios ou valores discrepantes ou inconsistentes. Os dados que compõe o *dataset* em questão são obtidos através de uma aplicação que gere as avaliações dos alunos, ou seja, que gere todo o processo de registo dos dados pelo que não permite a existência destes problemas nos dados brutos.

A integração dos dados é o processo de combinar dados derivados de várias fontes de dados num único conjunto de dados consistente sendo que há uma série de questões a serem consideradas durante a integração de dados. Contudo, os dados que são analisados são obtidos de uma única base de dados que armazena os registos obtidos de uma dada aplicação de gestão de avaliações pelo que segue padrões pré-definidos (Ahmed Mueen e Manzoor, 2016).

O objetivo da transformação dos dados é transformar os valores dos dados num determinado formato, escala ou unidade que seja mais adequado para análise, através de processos como a normalização, agregação ou generalização. Nesta etapa foi necessário codificar as variáveis categóricas em valores numéricos utilizando técnicas de codificação. Nos dados em estudo duas das variáveis são categóricas nominais, o tipo e subdomínio das questões, bem como uma variável categórica ordinal, o tipo da questão. As restantes variáveis são numéricas. Variáveis categóricas são variáveis que contêm etiquetas - *labels* em vez de valores numéricos. Existem dois tipos principais de variáveis categóricas: as nominais, que são variáveis que não estão relacionadas entre si em qualquer ordem, e ordinais, que são variáveis onde existe uma certa ordem entre elas. Existem várias as técnicas de codificação (Potdar et al., 2017), sendo que a *Label Encoding* (scikit learn, 2022c) é uma das soluções mais simples e comuns para transformar variáveis categóricas. Contudo, o *One-Hot Encoding* (scikit learn, 2022d) é a maneira mais comum e correta de lidar com dados categóricos não ordinais. Esta técnica

compara cada nível da variável categórica com um nível de referência fixo e transforma uma única variável com n observações e d valores distintos em d variáveis binárias com n observações cada uma. Cada observação indica a presença, com o valor 1, ou ausência, com o valor 0, da variável binária (Potdar et al., 2017). Também foi necessário fazer encoding da variável temporal, neste caso como se pretende preservar a natureza cíclica das suas entradas a abordagem foi decompor em outras variáveis com as características como o dia, mês, ano e dia da semana.

Após a transformação dos dados foi realizada a seleção de características - *feature selection* com o intuito de identificar os atributos que têm maior impacto no problema em questão. Embora o número de variáveis em estudo não seja elevado, algumas não estão relacionadas com o desempenho do aluno. A *feature selection* (Dash e Liu, 1997) é o processo em que são selecionados, automaticamente ou manualmente, os recursos que mais contribuem para a variável de previsão ou saída. Neste caso foi aplicado tanto a *feature selection* manual como a automática recorrendo à Matriz de Correlação. A correlação indica como os recursos estão relacionados entre si ou com a variável de saída. Como resultado da *feature selection*, na Tabela 3 são apresentadas as variáveis selecionadas. No problema em questão a análise realizada foi por domínio, subdomínio e aluno pelo que as variáveis em estudo diferem de acordo com a análise pretendida. Sendo assim, na análise por domínio e aluno todas as variáveis em análise são as indicadas. Porém, na análise por subdomínio esta mesma variável não estará em análise.

| Variável | Descrição |
|---------------------------|------------------------------|
| question.answering_time | Tempo de resposta da questão |
| question.subdomain | Subdomínio da questão |
| question.difficulty_level | Dificuldade da questão |
| question.type | Tipo da questão |
| timetag.day | Dia da resposta |
| timetag.month | Mês da resposta |
| timetag.year | Ano da resposta |
| timetag.weekday | Dia da semana |

Tabela 3: Resultado da *feature selection*

A análise complexa em grandes conjuntos de dados pode levar muito tempo ou mesmo ser inviável por isso a etapa final do pré-processamento dos dados é *data reduction*, ou seja, o processo de redução dos dados por meio de uma representação mais eficaz do conjunto de dados sem comprometer a integridade do dataset original. No caso que temos em mãos não foi necessário ser aplicado.

4.3 CLASSIFICAÇÃO DE DESEMPENHO

Num processo de classificação é constituído por duas partes, o treino e o teste. Os dados de teste – *test set* são utilizados para estimar a exatidão das regras de classificação. No treino, o modelo é construído utilizando parte dos dados conhecidos como conjunto de treino - *training set*, que conhecem todos os atributos, incluindo

as classes. Após a construção de um modelo, este é utilizado para definir uma etiqueta ou classe para um novo registo onde o atributo da classe é desconhecido (Ahmed Mueen e Manzoor, 2016). Foi utilizado um conjunto de dados para teste e treino dos modelos, utilizando a técnica *train-test split*, com divisão aleatória do mesmo: 70% das respostas foram utilizadas para treino e 30% para teste da qualidade dos modelos. Todos os modelos desenvolvidos utilizaram os mesmos dados de treino e de teste. O objetivo dos algoritmos de classificação é gerar resultados corretos e precisos. Existem numerosos métodos para a criação de conjuntos de classificadores, como DT, NN, Naïve Bayes e SVM. Foram selecionadas três técnicas de classificação com base na sua reputação na literatura recentemente publicada sobre mineração de dados e a sua superioridade em problemas deste tipo de previsão para analisar o problema em estudo.

A DT é um algoritmo de aprendizagem de máquina supervisionado de suporte à tomada de decisão cuja estrutura é em forma de árvore, representando conjuntos de decisões que por sua vez geram regras para a classificação de um conjunto de dados (Ahmed Mueen e Manzoor, 2016). Esta é uma técnica utilizada para classificação e também previsão de conjuntos de dados. A aprendizagem recorrendo a árvores de decisão é um método muito utilizado e poderoso na prospeção de dados. O objetivo é criar um modelo que preveja o valor de uma variável alvo com base em várias variáveis de entrada. Cada nodo interior da árvore corresponde a uma das variáveis de entrada e cada folha representa um valor da variável alvo dado os valores das variáveis de entrada representados pelo caminho desde a raiz até à folha (Kesavaraj e Sukumaran, 2013). Basicamente, como podemos ver na Figura 12, o conhecimento é representado numa DT por:

- **Nó raiz:** Nó raiz, que é o nó que executa a primeira divisão, que pode ser em dois ou mais sub-nós. Um sub-nó que se divide em outros sub-nós, denomina-se por nó de decisão. Quando removemos sub-nós de um nó de decisão, esse processo é chamado de poda - *pruning*. Com o método de *pruning*, os nós desnecessários da árvore de decisão podem ser removidos, para que se obtenham resultados mais refinados e se reduzam os erros.
- **Nós folhas/terminal:** Nó folha/terminal que são os nós que preveem o resultado. Quando o nó é dividido em sub-nós, este é chamado de nó pai e os sub-nós de nós filhos.
- **Ramos:** Ramos, que ligam os nós entre si e revelam o fluxo da resposta.

Por isso, o nó raiz é o ponto de partida da árvore e os nós folha/terminal contêm as perguntas ou critérios que desejamos obter resposta (da Rocha e de Sousa Junior, 2010).

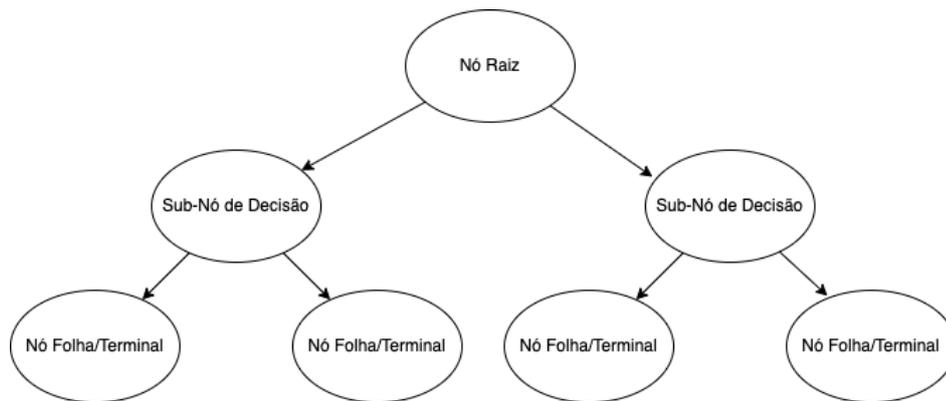


Figura 12: Estrutura de uma DT

As DT caracterizam-se pelos resultados derivados serem fáceis de interpretar e de compreender e, para além disso, em relação a outros métodos, não requerem um grande pré-processamento dos dados. Contudo, são mais propensas a sobreajuste – *overfitting* -, ou seja, ajustam-se muito bem ao conjunto de dados, mas mostram-se ineficazes para prever novos resultados. Este problema pode ser atenuado com a definição de uma profundidade máxima. Porém, isso pode introduzir um enviesamento e, portanto, comprometer a precisão dos resultados. As DT são muito sensíveis em relação aos atributos relevantes no conjunto de dados. Uma pequena alteração no *dataset* pode causar grandes alterações na estrutura da árvore. Além disso, tornam difícil a partição de atributos com dados contínuos, pelo que não são adequadas para conjuntos de dados com este tipo de atributos .

As RF são um algoritmo de aprendizagem supervisionada que cultiva muitas árvores de classificação. Na prática, produz uma coleção de árvores de decisão. Esta é uma técnica utilizada tanto em processo de classificação como em processo de regressão de conjuntos de dados. A principal diferença entre o algoritmo das DT e das RF é que o estabelecimento de nós de raiz e de nós de segregação, neste último, é feito de forma aleatória. Esta aleatoriedade provém do método *Bagging* que envolve a utilização de diferentes amostras de dados em vez de apenas uma amostra. As DT produzem resultados diferentes, dependendo dos dados de treino, que são posteriormente classificados e o mais elevado é selecionado como resultado final (Kha, 2020). Observando a Figura 13 verificamos que as RF dependem de várias RF e cada uma consiste em nós de decisão, nós de folha, e um nó de raiz. Os nós de folha de cada árvore são o resultado final produzido por essa DT específica. Neste simples classificador de árvores aleatórias, os nodos azuis são a previsão e a seleção da produção final, que segue o sistema de votação maioritária. O resultado escolhido pela maioria das DT torna-se o resultado final do sistema de floresta tropical.

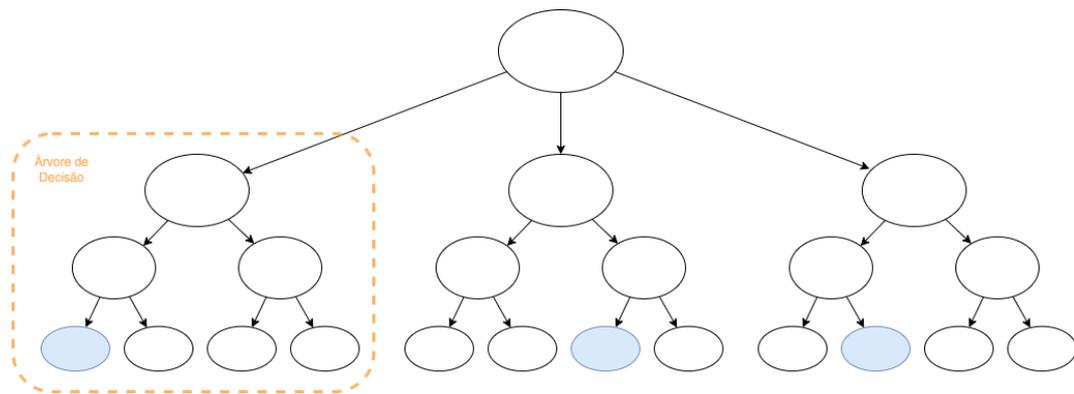


Figura 13: Estrutura de uma RF

As RF destacam-se pelo aumento da precisão e redução do problema de *overfitting* dos dados. Para além disso, o desempenho dos testes não diminui com o aumento do número de árvores, que é uma vantagem tendo em conta que a construção de inúmeros modelos de DT gera resultados mais exatos. Porém, este algoritmo utiliza mais recursos e consome mais tempo do que as DT, devido à sua aleatoriedade (Kha, 2020).

As SVM são modelos de aprendizagem supervisionados com algoritmos de aprendizagem associados, que analisam dados e reconhecem padrões. Como podemos constatar na Figura 14, a SVM básica toma um conjunto de dados de entrada e prevê, para cada uma das entradas, qual das duas classes forma a saída, tornando-a num classificador linear binário não-probabilístico. Os modelos SVM têm formas funcionais semelhantes a NN e funções de base radial, que são técnicas de DM populares. As SVM é um algoritmo de treino para a classificação da aprendizagem e regras de regressão, contudo a qualidade da generalização e facilidade de treino deste modelo está muito para além das capacidades dos mais tradicionais (Cortes e Vapnik, 1995).

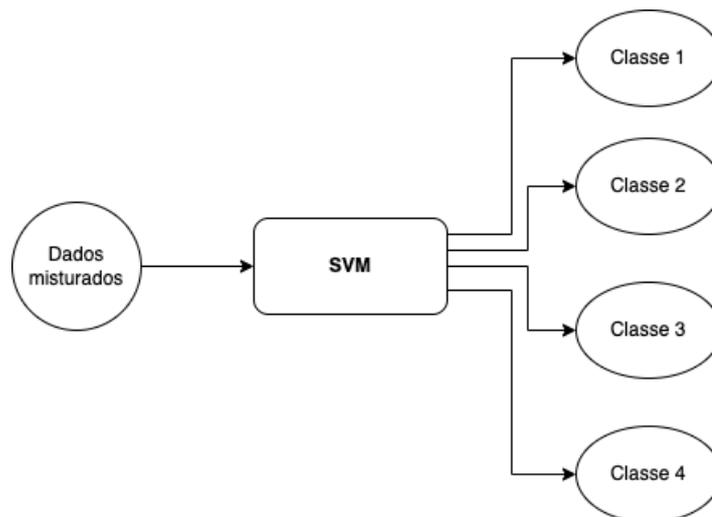


Figura 14: Processo do classificador SVM

Uma *SVM* generaliza muito bem as novas amostras quando os parâmetros são sabiamente escolhidos, sendo altamente eficaz mesmo com dados de alta dimensão, permitindo por vezes obter soluções únicas. Todavia, o resultado da *SVM* é dificilmente interpretável e conforme o tamanho do conjunto de dados em estudo vai aumentando o tempo necessário para fazer os cálculos associados cresce rapidamente e a interpretabilidade diminui (Karamizadeh et al., 2014).

Os três modelos escolhidos para aplicar no problema em questão foram as *DT*, as *RF* e as *SVM*, os resultados bem como as conclusões são apresentadas no Capítulo 5.

Dado que o objetivo é encontrar um modelo com um desempenho elevado ou que a taxa de erro seja menor, recorreremos à otimização dos hiper parâmetros - *Hyperparameter Optimization* (*HPO*) ou *tuning* que é o processo de afinação dos parâmetros enquanto construímos modelos de aprendizagem. Este processo apresenta vários casos de utilização importante, nomeadamente:

- Reduzir o esforço humano necessário para aplicar a aprendizagem mecânica;
- Melhorar o desempenho dos algoritmos de aprendizagem de máquinas;
- Melhorar a reprodutibilidade e a equidade dos estudos científicos (Bergstra e Bengio, 2012).

Existem, principalmente, dois tipos de métodos de otimização de *HPO*, em concreto a pesquisa manual e os métodos de pesquisa automática. A pesquisa manual experimenta os conjuntos de hiper parâmetros à mão, pelo que depende da intuição e da experiência do utilizador e à medida que o número de hiper parâmetros e a gama de valores aumenta, tornando-se, assim, bastante difícil de gerir (Wu et al., 2019).

Para superar os inconvenientes da pesquisa manual, foram propostos algoritmos de pesquisa automática, tais como a pesquisa em grade - *grid search*. O princípio deste algoritmo é a pesquisa exaustiva no qual há a formação de um modelo de aprendizagem com cada combinação de possíveis valores de hiper parâmetros no conjunto de formação e avalia o desempenho de acordo com uma métrica pré-definida. Contudo, a eficiência do algoritmo diminui rapidamente à medida que o número de hiper parâmetros é ajustado e a gama de valores dos hiper parâmetros aumenta (Wu et al., 2019).

Para resolver o problema do custo dispendioso do *grid search*, foi proposto o algoritmo de pesquisa aleatória - *random search*, uma vez que, tendo em conta que na maioria dos conjuntos de dados apenas alguns dos hiper parâmetros são realmente importantes, reduz a pesquisa a hiper parâmetros que não importam, e, finalmente, obtém-se a solução aproximada da função de otimização. Para tal, a pesquisa é aleatória através de combinações aleatórias de uma gama de valores (Wu et al., 2019).

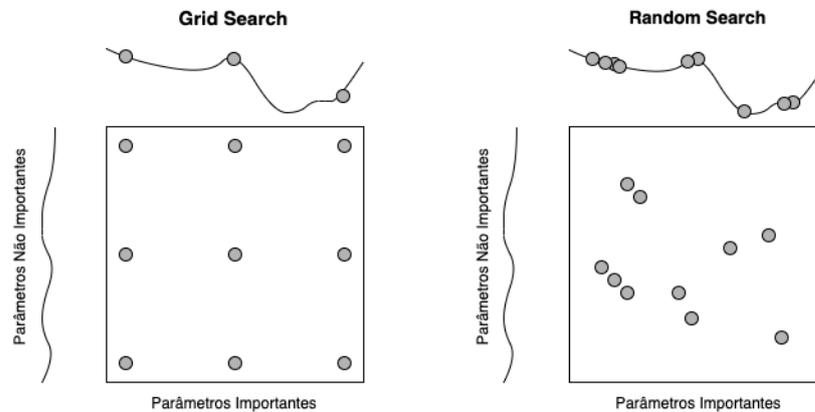


Figura 15: Comparação *grid search* e *random search* – figura adaptada de (Bergstra e Bengio, 2012)

Em comparação com o *grid search*, o *random search* é mais eficiente num espaço de alta dimensão. No entanto, não é fiável para treinar alguns modelos mais complexos. O *random search* torna a paralelização mais fácil e a atribuição de recursos mais flexível. Como se pode ver na Figura 15, comparando as duas abordagens, mostrando o espaço hiperparamétrico, a disposição aleatória mostra que o *random search* explora o espaço de forma mais ampla (Bergstra e Bengio, 2012), tendo isso em consideração, decidimos aplicar ambos os métodos de otimização e fazer a sua comparação no Capítulo 5.

No caso das DT, quando a árvore está a ser treinada, é possível que alguma folha tenha apenas uma observação associada, o que pode conduzir a uma situação de *overfitting*. Para controlar esse tipo de situação podemos usar os hiper parâmetros *min_samples_split* e *min_samples_leaf*. O primeiro destes parâmetros garante que um corte só é feito em nós com no mínimo de *min_samples_split* observações. No entanto, mesmo que esse número seja alto, é possível que alguma das folhas geradas por esse corte tenham um número muito baixo de observações. Para resolver este problema, podemos usar o *min_samples_leaf*, que determina o número mínimo de observações que cada folha deve ter. Assim, se um corte fosse criar uma folha com menos observações, este não seria realizado. Para além disso, podemos usar também *max_depth* para controlar a profundidade máxima da árvore. Outros hiper parâmetros importantes das DT são o *max_features*, que é o número de características a considerar quando se procura a melhor divisão, o *splitter*, que é a estratégia usada para escolher a divisão a cada nó, e o *criterion*, que é a função para medir a qualidade de uma divisão. Por último, o *random_state*, que permite selecionar de forma aleatória as características e amostras de dados. Sendo assim, para este caso, foram selecionados sete parâmetros no total e os seus respectivos valores para serem manipulados. As RF diferem dos hiper parâmetros indicados anteriormente uma vez que o parâmetro *splitter* não existe. Nas SVM existem três parâmetros que são habitualmente otimizados, nomeadamente, o *kernel*, que transforma as entradas de baixa dimensão numa dimensão superior, o C, que é o parâmetro de penalização, representando uma classificação ou termo errado, e, por último, o *gamma*, que define até que ponto chega a influência de um único exemplo de treino, ou seja, que o *gamma* alto considerará apenas pontos próximos ao hiper plano plausível enquanto que o *gamma* baixo considerará pontos a uma distância maior.

Neste trabalho foram aplicados os algoritmos de *grid search* e *random search* com *CV*. A *CV* é utilizada durante o treino do modelo, dividindo os dados de treino em duas partes, dados de treino e dados de validação. O tipo de *CV* mais popular é a *CV* em *K-Fold*, que é, basicamente, um processo iterativo que divide os dados de treino em *k* partições. Cada iteração mantém uma partição para teste e as restantes partições *k-1* para treino do modelo. A próxima iteração irá definir a partição seguinte como dados de teste e as *k-1* restante como dados do treino. E assim por diante. Em cada iteração, é registado o desempenho do modelo e, no final, dará a média de todo o desempenho. Facilmente se pode constatar que este processo é demorado (Arlot e Celisse, 2010).

Para a aplicação do *grid search* é também necessário definir parâmetros importantes. O estimador, que é o modelo em questão, os valores possíveis para os hiper parâmetros referidos anteriormente, o método para medir a performance do modelo, que, como no caso pretendemos fazer classificação, será a precisão – *accuracy* - e o número de *jobs* em paralelo a executar o algoritmo, que interfere diretamente com a velocidade do processo. Quando o valor é -1 são usados todos os cores disponíveis e o número de folds para a *CV*, que no caso é o padrão, 5 *folds*. No caso do *random search* utilizam-se a maior parte dos mesmos parâmetros, acrescentando, contudo, a definição do número de combinações para serem selecionadas aleatoriamente.

4.4 PREVISÃO DE DESEMPENHO

No processo de classificação, o modelo a derivar é o de prever a classe de objetos cuja etiqueta de classe é desconhecida. Geralmente, a classificação dos dados realiza-se num processo em duas fases, que são aprendizagem e uma etapa de classificação. Esta última é utilizada para prever etiquetas de classe para dados de formação (Mayilvaganan e Kalpanadevi, 2015). A classificação é uma técnica de prospeção de dados preditiva que faz previsão sobre valores de dados utilizando resultados conhecidos que foram encontrados a partir de dados diferentes. Os modelos preditivos têm o objetivo específico de nos permitir prever o valor desconhecido de uma variável de interesse, dados os valores conhecidos de outras variáveis (Pandey e Pal, 2011). As técnicas de previsão diferem das técnicas de classificação, uma vez que os registos são classificados de acordo com a previsão de comportamentos futuros ou dos valores estimados para o futuro. Os resultados obtidos apenas podem ser validados, esperando pelo acontecimento. Após a construção e a análise dos vários modelos de classificação indicados na secção anterior, foi possível prever o resultado da resposta a uma dada pergunta do questionário das matérias selecionadas e avaliadas, utilizando o modelo selecionado. Os modelos em questão encontram um padrão no conjunto de dados em análise, o que faz com que seja possível prever, para um dado conjunto de perguntas, o resultado que seria obtido.

Nesta dissertação pretendeu-se prever o desempenho dos alunos em situações de avaliação, uma vez que, para determinadas questões, pretende-se prever se o aluno irá acertar ou não a questão. Para tal teve-se em conta as diversas variáveis que pudessem influenciar o desempenho do aluno. Sendo assim, nesta fase o objetivo foi determinar, para um determinado período de tempo, de 5, 10 ou 15 dias, a taxa de acerto do aluno. Para tal, inicialmente foi selecionado um grupo de questões já respondidas pelo aluno, num determinado período de tempo. Este grupo de questões é constituído pelo número médio de questões respondidas diariamente neste período de análise, sendo estas questões selecionadas de forma aleatória. Após formado o grupo de questões

a analisar foi utilizado o modelo treinado e testado na fase anterior para prever se o aluno acerta ou erra cada uma das questões para cada dia. Após esta previsão foi, então, possível calcular a taxa de acerto, que resulta do cociente do número de questões corretas pelo número de questões respondidas.

4.5 CLUSTERING DE DESEMPENHO

Para lidar com a grande quantidade de dados gerados hoje em dia, um dos principais meios que temos para o fazer é classificá-los ou agrupá-los num conjunto de *clusters*. Estes sistemas de classificação podem ser supervisionados, caso tenham categorias supervisionadas discretas, que provêm de dados rotulados, ou não supervisionados, se tiverem categorias não supervisionadas, que provêm de dados não rotulados. No *clustering*, que é uma classificação não supervisionada, não existem dados rotulados. Basicamente, pretende-se separar este tipo de dados num conjunto discreto e estruturado com categorias ou agrupamentos. O principal objetivo da análise de *clusters* é dividir um grupo de objetos em subgrupos homogêneos, através de uma medida de semelhança escolhida (Xu e Li, 2005).

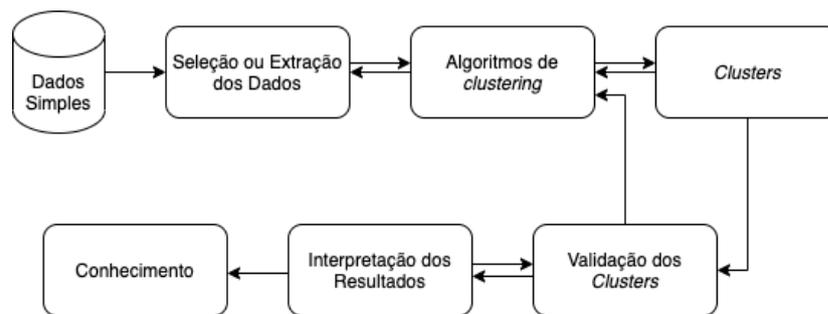


Figura 16: Ilustração do processo de *clustering* – figura adaptada de (Xu e Li, 2005)

Na Figura 16 está apresentada uma ilustração do processo de *clustering*. Nesse processo, podemos ver que, inicialmente, é feita a seleção ou extração dos dados. A seleção escolhe características distintivas de um conjunto de dados, enquanto que a extração utiliza algumas transformações para gerar características úteis e inovadoras a partir das originais. Ambas as operações são cruciais para a eficácia do processo, uma vez que podem diminuir a carga de trabalho dos algoritmos.

Após essas operações é feita a seleção do algoritmo de *clustering*, tarefa na qual é selecionada uma medida de proximidade e a construção de uma função critério. Os padrões são agrupados de acordo com as semelhanças entre si, mas a medida de proximidade afeta diretamente os resultados. Esta medida permite a construção de uma função de critério de *clustering* tornando este um problema de otimização. Desta forma, para a resolução de qualquer problema é fundamental investigar quais são as características, de modo a definir e implementar a estratégia mais apropriada. Dado um mesmo conjunto de dados o resultado do processo de *clustering* pode diferir, uma vez que este depende da abordagem escolhida, tanto na escolha do algoritmo como dos parâmetros. Na fase de validação é feita uma análise dos *clusters* obtidos, do seu número e do seu significado. Por fim é realizada a interpretação dos resultados. Neste processo pretende-se dar aos utilizadores uma

visão analítica dos dados originais de modo a resolver eficazmente problemas que possam ser encontrados (Xu e Li, 2005).

Os algoritmos de *clustering* podem ser categorizados em três tipos de estruturas diferentes. De referir, o particional, o hierárquico e o individual. Os algoritmos mais populares são os particionais e os hierárquicos. Os hierárquicos agrupam os dados com uma sequência de partições numa estrutura hierárquica, enquanto os particionais dividem diretamente os dados em algum número pré-definido de agrupamentos, sem uma estrutura hierárquica (Xu e Li, 2005).

Existem vários algoritmos de *clustering*. No entanto, estes algoritmos de agrupamento funcionam eficazmente em dados puramente numéricos ou puramente categóricos, pelo que a maioria deles têm um mau desempenho em tipos mistos, ou seja, com dados categóricos e numéricos.

O algoritmo *k-means* (scikit learn, 2022b) é um dos algoritmos mais populares que é utilizado em mineração de dados e no reconhecimento de padrões. Basicamente, porque é simples e fácil de implementar. Contudo, apesar da sua popularidade, este algoritmo também apresenta algumas desvantagens, já que apenas permite a utilização de dados numéricos, não podendo ser aplicado quando existem dados categóricos.

Para fazer o agrupamento de dados, uma alternativa ao *k-means* é a utilização do algoritmo *k-mode* (pypi, 2021). Este é o resultado da modificação do *k-means*, no qual foi substituída a função da distância euclidiana pela simples medida de dissimilaridade de correspondência, isso permite com que seja possível fazer o agrupamento de dados categóricos. Contudo, esta alternativa não permite realizar a análise de dados mistos (numéricos e categóricos) (Girsang et al., 2020).

Uma das soluções apresentadas na resolução de valores numéricos e categóricos para problemas de segmentação é redefinir a medida da distância e aplicar aos algoritmos já existentes. O *k-prototypes* (pypi, 2021) é um dos mais famosos métodos, que faz a integração dos métodos *k-means* e *k-modes* e utiliza uma medida de proximidade combinada para comparar as observações analisadas. O *k-prototypes* é um método conhecido e largamente utilizado para segmentação de dados numéricos e categóricos (Shihm et al., 2010).

A implementação *k-prototypes* é realizada em 4 fases distintas. Na primeira fase são lidos todos os parâmetros a partir dos dados, tais como a quantidade de dados, o número máximo de cada agrupamento, todos os atributos com tipo numérico e tipo categórico, bem como os nomes e tipos dos atributos. Na fase seguinte é feita a seleção inicial de *k* protótipos iniciais, para de seguida alocar cada objeto do conjunto de dados ao protótipo mais próximos, levando em consideração a medida de proximidade combinada. Após cada realocação, são atualizados os protótipos nos novos grupos, *clusters*. É analisado novamente a pertinência de cada objeto nos *clusters* formados, caso existam objetos que precisem de ser realocados são atualizados os protótipos nos *clusters* já formados. Este processo é repetido até que não exista nenhum objeto do conjunto de dados para ser alocado ou realocado (Girsang et al., 2020).

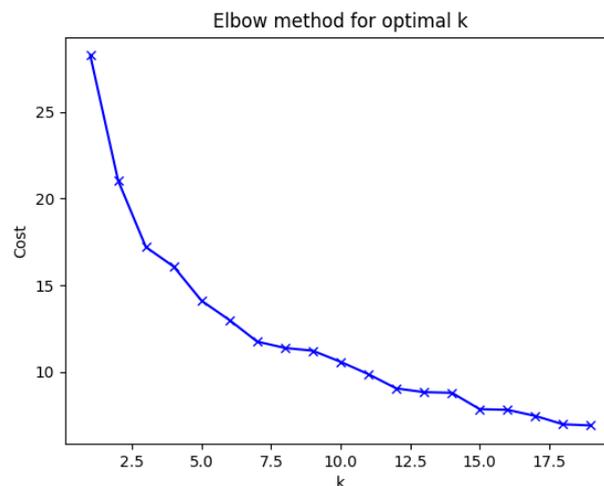
| Variável | Descrição | Tipo |
|---------------------------|------------------------------|---------|
| question.answering_time | Tempo de resposta da questão | Inteiro |
| question.subdomain | Subdomínio da questão | Objeto |
| question.difficulty_level | Dificuldade da questão | Inteiro |
| question.type | Tipo da questão | Objeto |

Tabela 4: Dados em análise no processo de *clustering*

O conjunto de dados em estudo é constituído por dados mistos - *mixed data types*, ou seja, com valores numéricos e categóricos em estudo, como podemos verificar na Tabela 4. Portanto, existiam duas soluções possíveis, a aplicação do algoritmo *k-means*, com *encoding* das variáveis categóricas, ou do algoritmo *k-prototypes*.

Nestes dois algoritmos, *k-means* e *k-prototypes*, é necessário indicar o valor de *k*, que representa o número de *clusters* nos dados. Existem alguns métodos que permitem identificar o número. Para esta investigação, foi aplicado o método *Elbow* (Bholowalia e Kumar, 2014), que é um método visual que tem como ideia base escolher um número de *clusters* de modo a que ao adicionar outro *cluster* não dá uma modelação muito melhor dos dados. A escolha do número de *clusters* é de extrema importância, uma vez que a percentagem de variação explicada pelos *clusters* é traçada em função do número de *clusters*. Os primeiros *clusters* acrescentam muita informação. Porém, a partir de uma dada altura, o ganho marginal cai drasticamente e é perceptível a formação de um ângulo no gráfico, sendo esse o valor ótimo de *k* (Bholowalia e Kumar, 2014).

O primeiro método aplicado sobre os dados em estudo no nosso caso foi o *k-prototypes*. Primeiramente foram normalizadas as variáveis contínuas do conjunto de dados, de modo a garantir que não seriam consideradas mais importantes que as categóricas. Recorrendo ao método *Elbow*, determinou-se o valor ótimo de *k* analisando o gráfico da Figura 17, para, de seguida, se aplicar o algoritmo *k-prototypes*. O algoritmo foi aplicado para três *clusters* distintos, sendo indicadas as colunas cujos dados são categóricos (colunas 1 e 3).

Figura 17: Identificação do *Elbow* para o processo *k-prototypes*

O outro método alternativo aplicado foi o *k-means*, com *One Hot Encoding* das variáveis categóricas. O *k* foi determinado através do mesmo método e o valor obtido também foi três, observando a Figura 18.

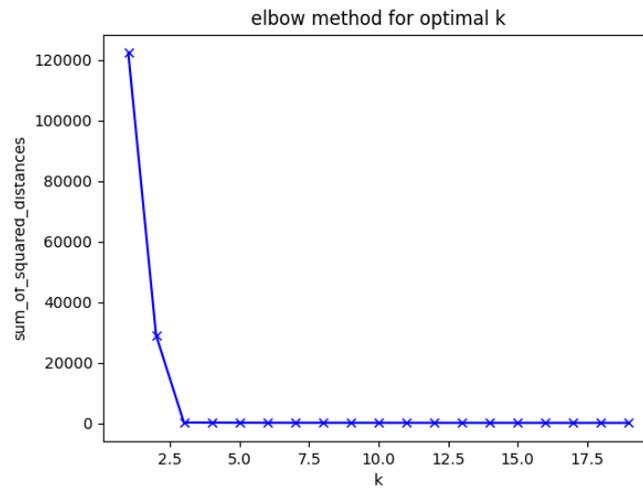


Figura 18: Definição do *Elbow* para o processo *k-means*

De seguida, no Capítulo 5, podemos ver os resultados obtidos, bem como a análise realizada e as conclusões que foram tiradas.

ANÁLISE DOS MODELOS IMPLEMENTADOS

5.1 CLASSIFICAÇÃO DO DESEMPENHO

Para a análise dos modelos discutidos e implementados foi utilizado um conjunto de dados - *dataset* constituído por dados sintetizados, gerados com um script em *Python* que construiu um conjunto de respostas bastante amplo, relativo a um período de 180 dias, tendo cada um dos dias cerca de 50 registos. O *dataset* final integrou cerca de 340 000 entradas. Cada modelo desenvolvido foi testado e treinado com os dados de acordo com o pretendido, isto porque se pretendia uma análise bastante detalhada, em particular, por aluno, por domínio e por subdomínio de estudo.

Foi implementado as *DT*, inicialmente sem qualquer tipo de otimização, recorrendo à classe *DecisionTreeClassifier*. Os dados usados para treinar e testar o modelo em questão são apenas constituídos por questões de um dado domínio. Na Figura 19 é possível ver a árvore resultante da aplicação do algoritmo de *DT* o *standard*, na qual podemos analisar a impureza *Gini* ([scikit learn, 2022a](#)), que é o valor do método de divisão do nó, em que quanto menor seja este valor, maior é a homogeneidade do nó, bem como o número total de amostras, os critérios de classificação e o número de amostras nós à esquerda e à direita do nó em questão. No início obtivemos 7000 amostras todas elas misturadas. Contudo, o valor da impureza de *Gini* já está próximo de zero, o valor já é relativamente baixo por isso já existem dados que estão bem separados. Escolhendo um dia menor que 19.5 passamos a ter um valor de impureza um pouco mais baixo, tendo já apenas 263 amostras. Neste caso foi verificado o valor do subdomínio, que tal como foi indicado em secções anteriores foi utilizado o método *One Hot Encoding* para processar estas variáveis categóricas pelo que se o valor do subdomínio for “Introdução aos Sistemas NoSQL” será 1 ou será 0 caso seja outro subdomínio. Desta forma, ao escolher um subdomínio “Introdução aos Sistemas NoSQL”, a condição é falsa e passamos a ter um valor de *Gini* bastante mais inferior e apenas 19 amostras, contudo se escolhermos um subdomínio que não seja este o valor aumenta um pouco e o número de amostras mantem-se elevado, 244. Esta análise pode ser assim feita ao longo de toda a árvore.

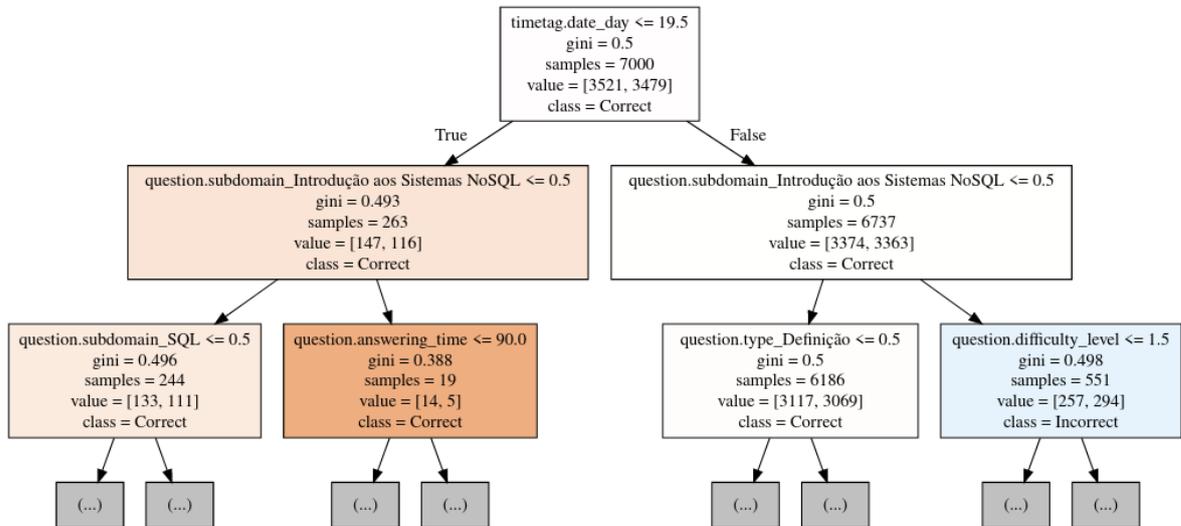


Figura 19: Resultado obtido: Árvore resultante da aplicação do modelo DT

Como se pode ver na Tabela 5 a *accuracy* obtida no treino foi de 77.3% e no teste de 56.0%, pelo que o modelo de base não é muito bom, uma vez que os valores obtidos são relativamente baixos. Desta forma, apesar de o modelo ter uma aprendizagem melhor com os dados de treino do que generalizar sobre novos dados (dados de teste), apresenta alguma dificuldade em generalizar. Contudo, não estamos perante um caso de *overfitting* nem *underfitting*. Este problema pode estar associado à própria natureza dos dados e dos parâmetros utilizados no algoritmo, que neste caso foram os *standards*. Por isso, recorreu-se aos métodos *grid search* e *random search* com *CV* para tentar melhorar o desempenho do nosso modelo.

Primeiro foi aplicado o *grid search*, com o objetivo de testar as várias combinações possíveis dos hiper parâmetros, selecionando aqueles que obtiveram menor erro. Os parâmetros selecionados permitiram definir valores como a profundidade máxima da árvore de 12 ou o número mínimo de amostras de cada folha e para fazer corte ser 4 e 10, respetivamente. A aplicação deste algoritmo afetou os resultados obtidos, tendo a *accuracy* do treino diminuído para 59% e 50% de teste.

De seguida foi aplicado o *random search* com os mesmos hiper parâmetros em análise. Como já era espectável neste caso, os valores ótimos mudaram. Por exemplo, a profundidade máxima da árvore foi agora de 18 e o número mínimo de observações de cada folha foi de 2. Com a aplicação deste algoritmo os valores da *accuracy* mantiveram-se semelhantes aos obtidos anteriormente, sendo 58% no treino e no teste 50%. O tempo de execução deste algoritmo foi muito menor do que o anterior, o *grid search*.

De seguida foi analisado o mesmo modelo com os dados explorados por subdomínio de conhecimento. Neste caso utilizámos o subdomínio “Álgebra Relacional” no processo de análise. Foi analisado o modelo base e também o modelo com a otimização dos hiper parâmetros. Os resultados obtidos foram muito semelhantes aos obtidos anteriormente, tal como podemos observar na Tabela 5.

Por último, foram analisados os dados por aluno, tendo-se selecionado, a título de exemplo, o aluno ‘u001 - Ana Castro e Cunha’. Neste caso a *accuracy* do treino obtida no modelo base foi muito elevada 96% e a de teste

manteve-se semelhante, o que nos permite observar um caso de *overfitting* no qual o modelo tem dificuldade em generalizar. Através da aplicação dos algoritmos *grid search* e do *random search* tornou-se possível melhorar a performance do algoritmo de *DT* aplicado anteriormente, não se tendo verificado *overfitting*.

| Em análise | Modelo | <i>Precision</i> | <i>Recall</i> | <i>Accuracy</i> treino | <i>Accuracy</i> teste |
|------------|-------------------------|------------------|---------------|------------------------|-----------------------|
| Domínio | <i>DT</i> | 0.51 | 0.60 | 0.773 | 0.56 |
| | <i>DT Grid Search</i> | 0.51 | 0.55 | 0.59 | 0.50 |
| | <i>DT Random Search</i> | 0.50 | 0.61 | 0.58 | 0.50 |
| Subdomínio | <i>DT</i> | 0.52 | 0.60 | 0.77 | 0.51 |
| | <i>DT Grid Search</i> | 0.51 | 0.71 | 0.51 | 0.50 |
| | <i>DT Random Search</i> | 0.52 | 0.58 | 0.52 | 0.57 |
| Aluno | <i>DT</i> | 0.54 | 0.53 | 0.96 | 0.52 |
| | <i>DT Grid Search</i> | 0.51 | 0.58 | 0.69 | 0.53 |
| | <i>DT Random Search</i> | 0.50 | 0.51 | 0.62 | 0.53 |

Tabela 5: Síntese dos resultados obtidos com o modelo *DT*

Analisando os dados obtidos na fase de treino e de teste, apresentados na Tabela 5, o modelo *DT* apresenta resultados abaixo do esperado com uma *accuracy* de treino a rondar os 70% e uma *accuracy* de teste de apenas 50%. Para além disso, a otimização dos hiper parâmetros com *CV* habitualmente permite obter melhores resultados pelo que era expectável uma melhoria da *accuracy*, o que não se verificou. Existem diversas justificações para a obtenção de resultados como estes, contudo tendo em conta que o processo de implementação decorreu de forma normal e o processo adotado foi o correto, a justificação mais plausível para este caso em específico será uma limitação dos dados ou então do modelo em análise. Foram implementados e analisados outros modelos de modo a comparar os resultados obtidos e, desta forma, perceber o modelo mais adequado para o caso em estudo.

O modelo analisado posteriormente foi as *RF*, primeiro o modelo base através da classe *RandomForestClassifier* com os parâmetros *standard*. Também foi feita a análise dos resultados por domínio, subdomínio e aluno uma vez que pretendemos que o modelo fosse adequado para os requisitos inicialmente definidos. Na Figura 20 podemos ver a *DT* resultante do primeiro estimador do modelo *RF*. A análise desta árvore difere da obtida no modelo anterior na Figura 19 porque o número de amostras é bastante menor uma vez que este modelo é a combinação de várias *DT*.

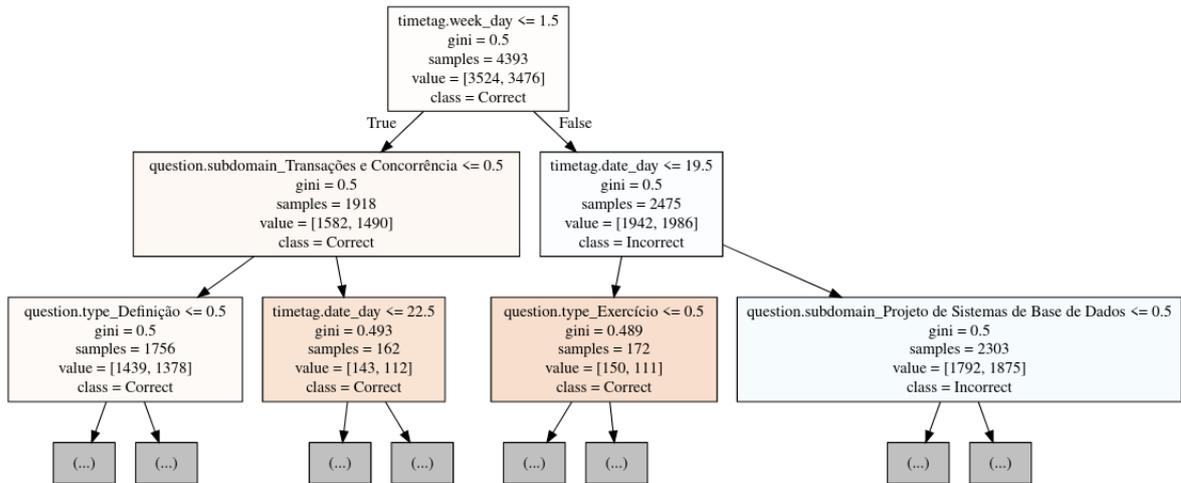


Figura 20: Resultado obtido: Árvore resultante da aplicação do modelo RF - 1º estimador

Na Tabela 6, podemos observar que o modelo RF sem qualquer tipo de otimização se comportou de forma semelhante ao modelo DT analisado anteriormente, como tal voltou-se a aplicar o *grid search* para otimizar os hiper parâmetros. Pela análise dos resultados obtidos (Tabela 6) verificou-se um aumento da performance em relação ao modelo DT, tanto na análise por domínio como por aluno. Além disso, deixou também de existir *overfitting* neste último caso. Por fim, foi ainda aplicado o *random search*. Verificámos que a performance melhorou em todos os casos e que o tempo de execução do *random search* foi bastante inferior.

| Em análise | Modelo | Precison | Recall | Accuracy treino | Accuracy teste |
|------------|------------------|----------|--------|-----------------|----------------|
| Domínio | RF | 0.51 | 0.50 | 0.72 | 0.50 |
| | RF Grid Search | 0.51 | 0.55 | 0.58 | 0.50 |
| | RF Random Search | 0.50 | 0.55 | 0.65 | 0.50 |
| Subdomínio | RF | 0.49 | 0.51 | 0.76 | 0.50 |
| | RF Grid Search | 0.51 | 0.54 | 0.59 | 0.50 |
| | RF Random Search | 0.51 | 0.57 | 0.67 | 0.51 |
| Aluno | RF | 0.52 | 0.49 | 0.95 | 0.52 |
| | RF Grid Search | 0.48 | 0.68 | 0.58 | 0.51 |
| | RF Random Search | 0.51 | 0.50 | 0.74 | 0.5 |

Tabela 6: Síntese das métricas obtidas com o modelo RF

Analisando os dados obtidos na fase de treino e de teste, apresentados na Tabela 6, o modelo RF também apresenta resultados abaixo do esperado com uma *accuracy* de treino a rondar os 75% e uma *accuracy* de teste de apenas 50%. Neste modelo mesmo com a otimização dos hiper parâmetros o comportamento foi semelhante ao obtido no modelo DT analisado anteriormente.

O último modelo que foi treinado e testado foram as SVM tanto o modelo base como com a otimização dos hiper parâmetros. Os resultados obtidos foram bastante baixos relativamente aos obtidos nos modelos anteriores, com uma *accuracy* de treino e teste a rondar os 50%, como podemos observar Tabela 6.

| Em análise | Modelo | <i>Precision</i> | <i>Recall</i> | <i>Accuracy</i> treino | <i>Accuracy</i> teste |
|------------|-------------------|------------------|---------------|------------------------|-----------------------|
| Domínio | SVM | 0.51 | 0.42 | 0.51 | 0.49 |
| | SVM Grid Search | 0.49 | 0.57 | 0.60 | 0.50 |
| | SVM Random Search | 0.47 | 0.53 | 0.60 | 0.50 |
| Subdomínio | SVM | 0.49 | 0.48 | 0.53 | 0.50 |
| | SVM Grid Search | 0.51 | 0.48 | 0.62 | 0.50 |
| | SVM Random Search | 0.49 | 0.50 | 0.51 | 0.50 |
| Aluno | SVM | 0.49 | 0.51 | 0.52 | 0.50 |
| | SVM Grid Search | 0.50 | 0.55 | 0.53 | 0.50 |
| | SVM Random Search | 0.49 | 0.51 | 0.54 | 0.51 |

Tabela 7: Síntese dos resultados obtidos com o modelo SVM

Terminada a análise de todos os modelos, verificámos que nenhum deles apresentou uma performance relativamente boa, uma vez que a *accuracy* de treino rondou os 60/70% e a de teste manteve-se apenas pelos 50%, mesmo tendo-se feito a otimização dos hiper parâmetros e com a utilização de CV. Contudo, foram resolvidos alguns problemas de *overfitting* nos modelos DT e RF.

Tendo sido feita uma análise de três modelos diferentes, o que poderemos concluir é que o problema em questão será derivado dos dados em análise, ou seja, a sua natureza e não o modelo em questão. Uma vez que os dados foram gerados de forma aleatória por um *script*, o modelo conseguiu aprender relativamente bem as relações existentes nos dados de treino, mas apresenta dificuldades a generalizar dado que apresenta dificuldades a encontrar os mesmos padrões nos dados de teste. Desta forma, apesar dos resultados obtidos não terem ido de encontro às expectativas, foi selecionado o modelo RF, uma vez que foi o modelo que apresentou melhores resultados e que se mantiveram estáveis nos vários casos analisados. A otimização selecionada foi com *random search* uma vez que os resultados obtidos com ambos os métodos de otimização foi semelhante, mas o tempo de execução do *grid search* é extremamente elevado em comparação com o *random search*.

5.2 PREVISÃO

A análise de uma série temporal pretende, por norma, alcançar dois objetivos: a modelação e a previsão dos valores futuros. O primeiro objetivo consiste em encontrar um modelo que tenha em conta as relações existentes entre as observações, permitindo a descrição da série temporal e o segundo objetivo diz respeito à previsão de valores futuros.

A previsão é o processo de fazer previsões sobre o futuro com base em dados do passado e tem um grau de risco e incerteza ligado a ela.

A aplicação do modelo selecionado anteriormente, RF, tornou possível analisar qual seria o comportamento do(s) aluno(s) no futuro, tendo em conta os dados que temos até ao momento, ou seja, da série temporal. Esta série temporal que será analisada pode ser definida pelo utilizador, podendo ser de dias, semanas, meses ou até mesmo anos. Desta forma, é agora possível observar o desempenho de um aluno nos 5, 10 ou 15 dias seguintes ao último dia com dados de questões respondidas no *quiz*. Para cada um dos dias é possível observar

a taxa de acerto, ou seja, o número de questões acertadas, tendo em conta o número de respostas que foram dadas. Na Figura 21 podemos analisar o comportamento de um determinado aluno, ou seja, analisar qual foi o desempenho do aluno num determinado período de tempo (19 de novembro a 24 de novembro) bem como o desempenho que é previsto obter nos dias seguintes (5 dias). Apesar da *accuracy* obtida pelo modelo RF desenvolvido ter sido relativamente baixa (cerca de 75%) verificámos que as previsões obtidas habitualmente mantem a tendência do desempenho que o aluno apresentava recentemente, não sendo verificadas grandes irregularidades.

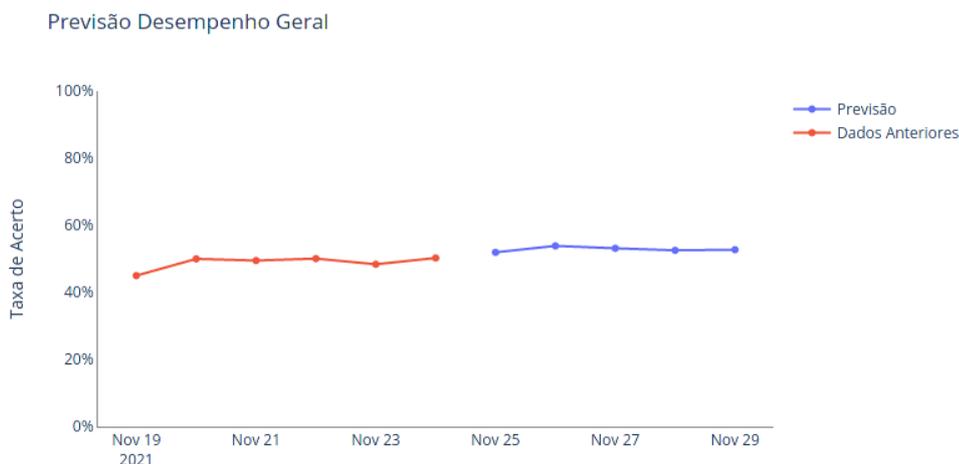


Figura 21: Previsão do desempenho (taxa de acerto) de um aluno

São vários os tipos de previsão disponíveis, para além do desempenho de um determinado aluno (Figura 21), é possível obter a previsão do desempenho por domínio e subdomínio.

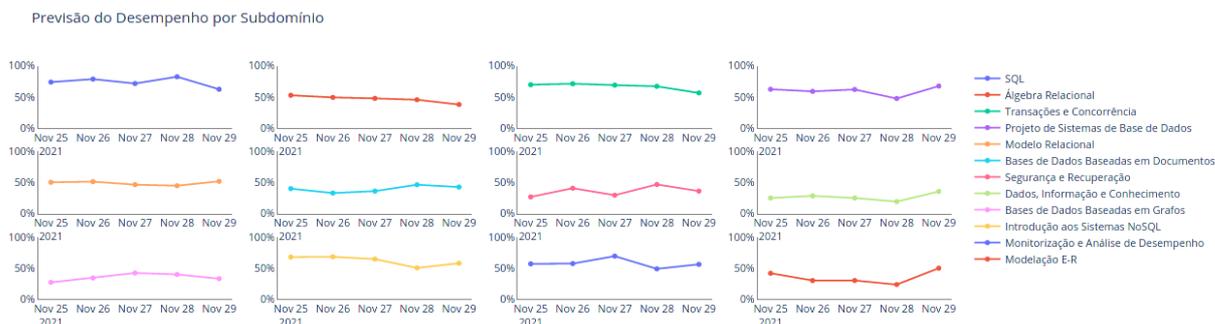


Figura 22: Previsão do desempenho (taxa de acerto) por subdomínio

A previsão apresentada anteriormente (Figura 21) é uma previsão do desempenho em geral, mas também foi possível fazer previsões mais específicas como é o caso do desempenho por subdomínios (Figura 22). Analisando as previsões obtidas verificamos que estas habitualmente mantêm-se semelhantes ao longo dos dias, não existindo casos que evidenciem previsões sem sentido.

As previsões obtidas foram de encontro aos objetivos definidos inicialmente, uma vez que permitem uma análise clara e explícita do desempenho dos alunos. Os resultados foram bons dado que mantêm a tendência de desempenho obtido pelo aluno ao longo do tempo e não foram encontradas previsões que realçassem grandes problemas no modelo desenvolvido. Contudo, a previsão tem um grau de risco e incerteza associado e o modelo desenvolvido apresenta uma *accuracy* de apenas de cerca de 75%.

5.3 CLUSTERING DO DESEMPENHO

Nesta fase o objetivo é criar grupos homogêneos, identificando os *clusters*. Para tal nesta dissertação foram implementados dois algoritmos distintos de *clustering*, o algoritmo *k-prototypes*, que é uma combinação dos algoritmos *k-means* e *k-modes*, e o algoritmo *k-means* com *One Hot encoding*.

Nas Figura 23 e Figura 24, podemos observar os gráficos obtidos com a aplicação do algoritmo *k-prototypes*. Como podemos observar existem 3 diferentes *clusters*. Na Figura 23 temos dois gráficos, com o tempo de resposta e a dificuldade da questão. Existem 3 *clusters* distintos, o primeiro *cluster*, *cluster 0*, é constituído por questões com maior dificuldade (dificuldade 2 a 5) e o tempo de resposta mais baixo a médio. O segundo *cluster*, *cluster 1*, o algoritmo agrupou questões com menos dificuldade (dificuldade entre 1 e 4) e um tempo de resposta baixo a elevado. Por fim, o terceiro *cluster* é constituído por todos os níveis de dificuldade (dificuldade 1 a 5) e o tempo de resposta médio a alto. Desta forma, podemos concluir que o tempo de resposta não está diretamente relacionada a dificuldade das questões.

Na Figura 24 estão apresentados gráficos com os subdomínios e tipos das questões. Analisando os resultados obtidos, verificamos que todos os *clusters* são constituídos por todos os subdomínios e tipos de questões.

Normalmente, numa análise dos resultados obtidos no *clustering*, é fácil e intuitivo identificar os *clusters*. Porém, neste caso isso não se verificou neste caso, tendo sido difícil tirar conclusões dos resultados obtidos. O *dataset* em estudo é constituído por dados sintetizados e a geração aleatória dos dados não varia muito o que pode justificar a distribuição dos dados e a dificuldade de agregação dos mesmos.

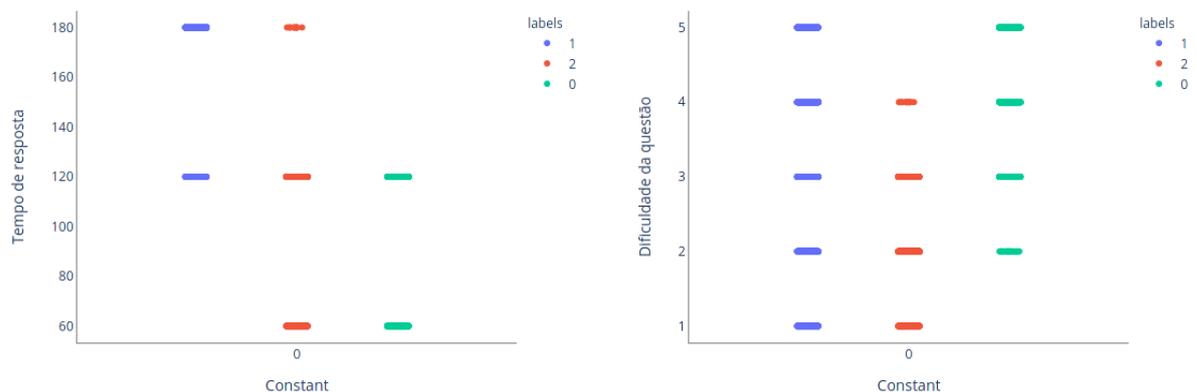


Figura 23: Resultados obtidos: *clustering* com *k-prototypes*

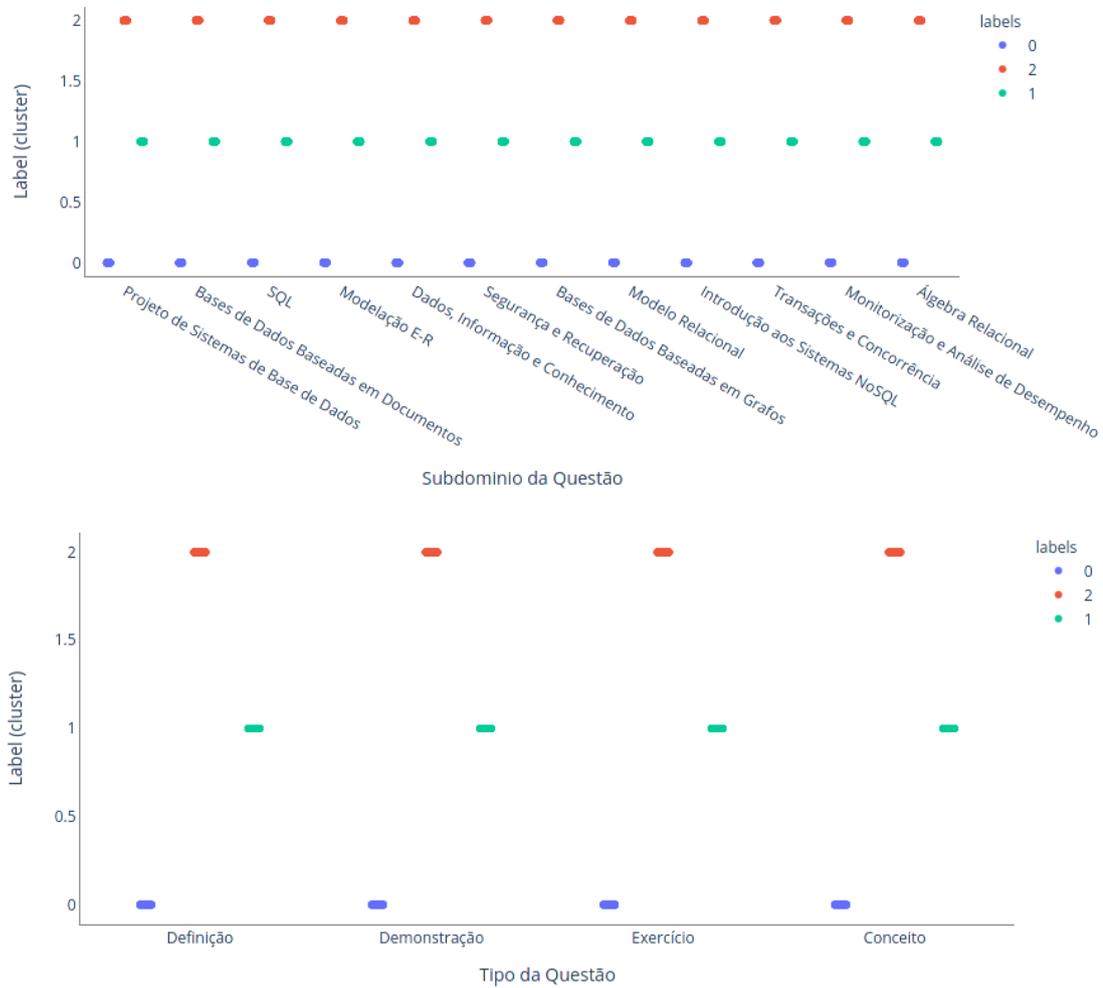


Figura 24: Resultados obtidos: *clustering* com *k-prototypes*

De seguida foi aplicado o algoritmo *k-means* com *One Hot Encoding*, tendo sido os resultados obtidos os apresentados na Figura 25. Como podemos observar foram obtidos também 3 *clusters*, tal como no algoritmo apresentado anteriormente. Contudo, neste caso todos os *clusters* apresentados são constituídos por questões dos vários níveis de dificuldade. No primeiro *cluster*, *cluster 0*, as questões são respondidas em tempo médio; no segundo *cluster*, *cluster 1*, o tempo é baixo; e no último *cluster* o tempo de resposta é elevado. Os resultados obtidos nos gráficos do subdomínio e tipo da questão são exatamente iguais aos obtidos pelos algoritmos *k-prototypes* (Figura 24).

Com a aplicação deste algoritmo tornou-se ainda mais difícil identificar os *clusters* do que com o algoritmo anterior e conseguir interpretar os resultados, não sendo possível concluir qualquer tipo de relação entre as variáveis em estudo.

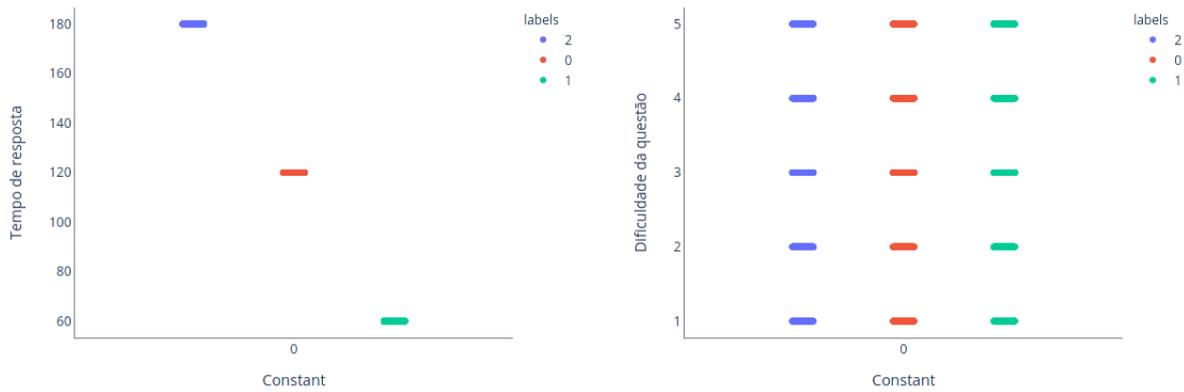


Figura 25: Resultados obtidos: *clustering* com *k-means*

Aqui, esperava-se que os resultados obtidos com os dois algoritmos fossem significativamente diferentes, uma vez que no *k-prototypes* os atributos categóricos e os numéricos têm o mesmo peso enquanto que o *k-means* pesa mais as variáveis categóricas devido ao método de *encoding* utilizado. Porém, os gráficos obtidos tanto do subdomínio como do tipo das questões, ambas variáveis categóricas, foi igual em ambos os algoritmos.

Com nenhum dos métodos utilizados foi possível encontrar uma relação direta entre as variáveis em estudo e identificar facilmente os *clusters*. Os resultados obtidos evidenciaram problemas no conjunto de dados, utilizado tanto para teste e treino. É perceptível através da observação dos gráficos que os dados são sintetizados, tendo sido gerados de forma aleatória.

VISUALIZAÇÃO E ANÁLISE DE DADOS DE DESEMPENHO

6.1 DASHBOARDS

A visualização de dados permite converter o que é abstrato e complexo em concreto e simples (Park e Jo, 2015). Ferramentas de visualização adequadas, tais como os *dashboards*, fornecem uma forma útil de visualizar dados e informações, da perspectiva de quem toma decisões (Teasley, 2017). O principal objetivo destas ferramentas é que seja possível uma percepção clara da informação, possibilitando identificar novas ligações ou padrões entre os dados, ter uma compreensão mais aprofundada dos dados ou gerar novas maneiras de pensar.

Estes painéis de visualização de dados são considerados mais do que uma simples coleção de indicadores e de gráficos, com formas e interfaces variadas, são capazes de fornecerem visualmente os elementos de dados necessários para responder às mais diversas perguntas, de acordo com as dimensões de análise que neles possam estar refletidas (Bolos et al., 2014a).

Por norma, os resultados exibidos pelos *dashboards* refletem a combinação de várias métricas de análise do domínio em questão, revelando tendências, reportando graficamente alarmes ou mesmo grelhas de resultados muito sofisticadas, como as *pivot tables*. Todos os elementos que um *dashboard* possa conter (e.g. valores, aspetos gráficos, grelhas ou indicadores de desempenho) podem ser configurados e geridos ao longo do tempo de forma acessível e flexível (Muntean et al., 2010).

Os *dashboards* são um dos casos mais comuns de utilização para visualização de dados, e a sua conceção e contextos de utilização são consideravelmente diferentes das ferramentas de visualização e análise exploratória de dados. Em comparação, os painéis interativos para a visualização de dados reúnem inúmeros desafios, como permitir a leitura de relance, pontos de vista coordenados e rastreamento de dados. Existem duas grandes perspectivas de *design* diferentes dos *dashboards*, os visuais em que os dados são estruturados em gráficos simples ou mais complexos, e os funcionais constituídos por um ecrã interativo que permite a monitorização em tempo real de dados atualizados dinamicamente (Sarikaya et al., 2019).

O desenho de *dashboards* não é uma tarefa simples, a maioria falha na comunicação eficiente e eficaz, devido a escolhas não certas dos componentes de visualização da informação, detalhe ou precisão excessiva, má codificação de dados quantitativos ou qualitativos, medidas apresentadas de forma errada, entre outros. Para não serem cometidos erros e sejam feitas representações contextualmente adequadas e com uma linguagem

visual adequada, existe muita literatura que descreve as técnicas e diretrizes que devem ser seguidas (Few, 2006a).

6.2 TÉCNICAS E COMPONENTES DE VISUALIZAÇÃO

Os *dashboards* são uma ferramenta de visualização de dados bastante adequada para auxiliar na tomada de decisões. Os painéis de visualização seguem diversas normas de modo que seja possível monitorizar as informações apresentadas de relance. Para tal, as informações mais importantes devem ser consistentes e destacar-se das restantes permitindo assim identificar informação relevante para a tomada de decisão. Recorrendo-se a tecnologias de visualização de dados é possível aumentar a perceção das informações (Park e Jo, 2015). Um *dashboard* com o intuito de permitir a monitorização de desempenho, deve apresentar uma visão dinâmica e única dos elementos que se pretendem analisar através da utilização de componentes visuais específicos, que permitam destacar dados e exceções que requerem uma atenção particular. A escolha das técnicas de visualização adequadas para cada tipo de informação que se pretende exibir é uma das tarefas mais importantes e desafiantes para o desenvolvimento de um *dashboard* efetivo e eficaz (Few, 2006b).

Segundo Few (2006), de toda a informação visual que nos chega ao cérebro, apenas uma pequena parte é realmente percebida e apenas uma parte ainda mais pequena é guardada na memória de curto prazo. Por esta razão não se deve sobrecarregar o utilizador de informação. A complexidade das visualizações num *dashboard* pode limitar a sua compreensibilidade, pelo que os componentes e técnicas utilizadas devem ser bem escolhidas. Os elementos de visualização podem ser básicos, como os diversos tipos de gráficos, de barras ou linhas, intermédios, que já incluem eixos duplos combinados, diagramas de dispersão e medidas cumulativas, ou avançados, quando se apresentam de visualizações de rede, barras ou intervalos de erro, *scatterplots* conectados, ou outras visualizações personalizadas (Sarıkaya et al., 2019). A maioria dos componentes de visualização enquadra-se nas tabelas ou gráficos básicos, uma vez que funcionam bem para transmitir informações básicas. Porém, a sua eficácia torna-se limitada para indicadores mais complexos, uma vez que estes são componentes mais adequados para cenários cujo volume de dados não é grande. A tabela é um elemento visual utilizado frequentemente para apresentar uma lista de informações quantitativas, permitindo a procura e comparação de valores. Contudo, para que exista uma boa interpretação da informação é importante que seja complementada com outras técnicas visuais.

No universo da visualização de dados, existem gráficos de barras, circulares, histogramas, gráficos de pontos, de caixa, de dispersão, de rosas, de mosaico e muitos outros tipos. A escolha da técnica de visualização mais adequada para cada tipo de informação é uma das tarefas mais importantes e desafiantes para o desenvolvimento de um *dashboard* efetivo e eficaz. A escolha depende do tipo de dados a serem exibidos, da quantidade de variáveis envolvidas, e daquilo que deve ser mostrado (houh Chen et al., 2008). Na Figura 26 é apresentado um esquema para auxiliar na escolha da visualização adequada para cada tipo de dado a ser exibido. De modo a selecionar o componente adequado é necessário fazer uma séria de questões. Primeiro, se é pretendido visualizar valores únicos ou valores múltiplos. Para exibir múltiplos valores, existem quatro categorias de elementos:

- **Relação** - como múltiplas variáveis independentes que se relacionam umas com as outras.
- **Comparação** - como dois ou mais conjuntos de dados se comparam um com o outro.
- **Composição** - como um conjunto de dados é composto por divisões menores.
- **Distribuição** - como os diferentes conjuntos de dados estão espalhados por uma população ou outra distribuição.

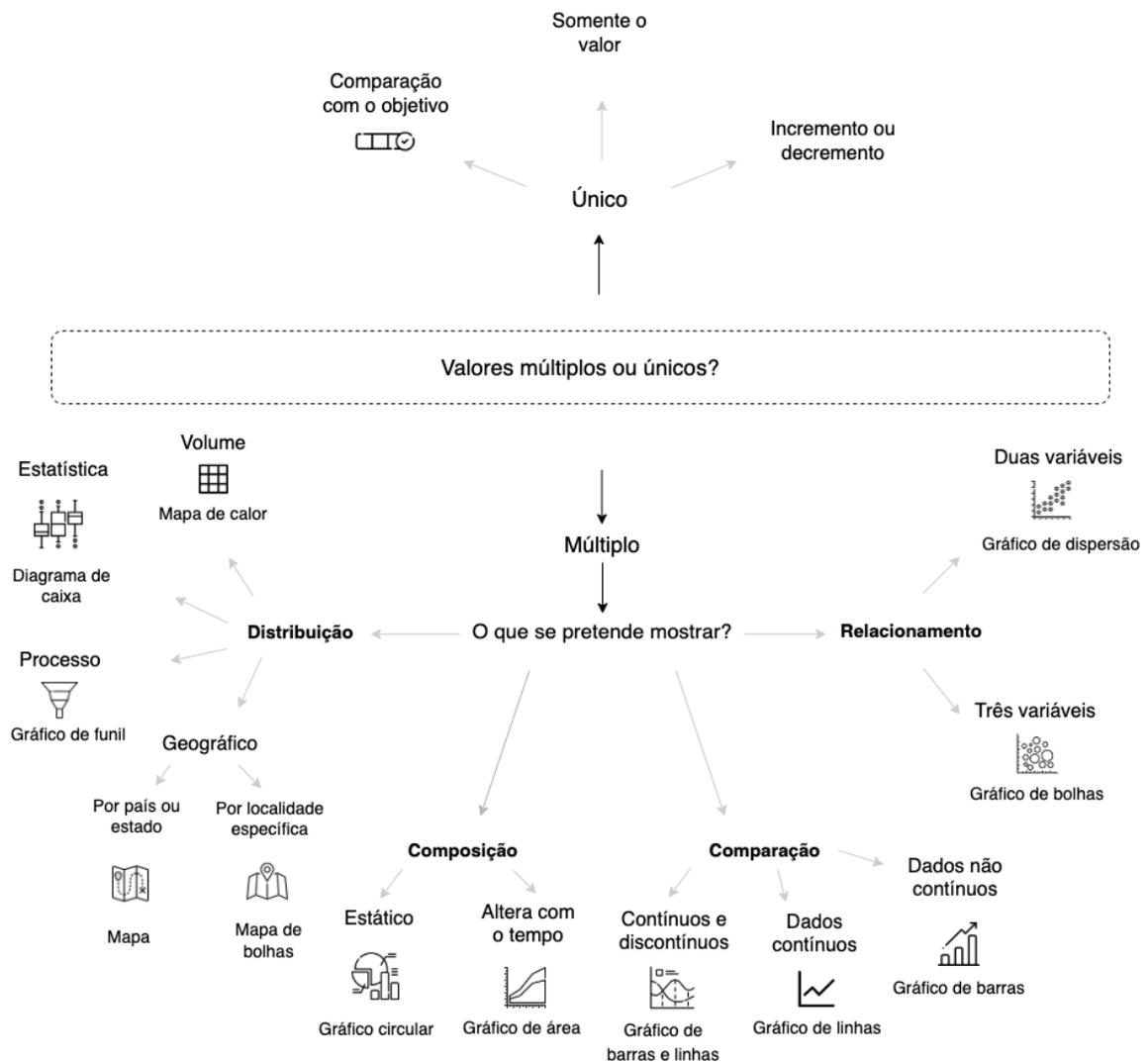


Figura 26: Diagrama de escolha de tipo de visualização – figura adaptada de (Tim, 2019)

Esta etapa deve ser concluída com cuidado, porque um gráfico tem ou não sucesso enquanto ferramenta de monitorização e de análise, dependendo do assunto, do contexto, do que se pretende mostrar e das considerações estéticas (houh Chen et al., 2008).

A representação orientada para a decisão dos resultados é um dos objetivos deste trabalho de dissertação, pelo que a informação complexa, não estruturada e proveniente de várias fontes foi transformada de modo a ser transmitida de forma resumida, clara e rápida nos *dashboards*. Desta forma, a construção do painel de visualização de dados é um processo direto, a partir do momento em que são recolhidos os dados e os indicadores calculados (Few, 2006b).

6.3 MODELOS E COMPONENTES IMPLEMENTADOS

O principal objetivo do desenvolvimento de *dashboards* para o sistema com que estamos a trabalhar é monitorizar o desempenho dos alunos fornecendo *feedback* sobre as atividades de aprendizagem e a evolução dos estudantes, identificar quem pode estar em risco, e fornecer informações sobre a evolução das interações dos estudantes na aprendizagem.

O nosso sistema tem como objetivo principal permitir analisar o desempenho dos alunos ao longo do tempo, para tal foi analisado um conjunto de dados com as respostas dos alunos a *quizzes* constituídos por questões de diversas matérias de uma determinada unidade curricular. De modo a auxiliar na análise e tomada de decisão recorreremos aos *dashboards*. Através destes painéis de visualização é possível observar padrões e sequências, sendo inferidas tendências no desempenho dos alunos. Desta forma, *dashboards* que exibam os resultados obtidos pelos alunos, mas também as previsões de desempenhos são importantes, bem como *dashboards* que permitem analisar o conjunto de dados em questão, através da observação e análise dos modelos de classificação e de *clustering*.

Para a monitorização do desempenho dos alunos, um dos indicadores chave é como os dados podem ser organizados em intervalos de dados ou classes. Neste trabalho de dissertação foi concluído no Capítulo 5, que o modelo mais apropriado para a classificação do conjunto de dados foram as RF. O resultado deste processo são várias DT, pelo que o mais correto é apresentar uma parte da árvore resultante de um dos estimadores, de modo a perceber como os dados estão a ser organizados. Para além disso, outras métricas obtidas como a *accuracy* do modelo aplicado ou a profundidade das árvores devem ser apresentadas, estas informações são indicadas para tabelas, permitindo visualizar informação detalhadamente.

Outro dos indicadores é a taxa de acerto em questões respondidas por um aluno, ou por vários alunos, num determinado domínio ou subdomínio, ou seja, a matéria a que a questão corresponde, num determinado período de tempo bem como a previsão da mesma. As previsões ao longo do tempo percebem-se melhor na forma gráfica, uma vez que os gráficos também reduzem a sobrecarga de informação quando comparados com a informação tabular.

Analisando a Figura 26, constatamos que pretendemos fazer uma comparação de dados contínuos, uma vez que pretendemos analisar o relacionamento existente num conjunto de dados contínuos ao longo do tempo. Neste caso, o gráfico indicado é o de linhas. Desta forma, torna-se possível revelar tendências ou mudanças ao longo do tempo.

O último indicador é a relação das variáveis em estudo, ou seja, perceber qual a distribuição dos dados, se existem valores atípicos ou algum padrão através da segmentação do desempenho. Os gráficos de relaciona-

mentos são usados para mostrar a conexão ou correlação entre duas ou mais variáveis. Sendo assim, para este indicador recorreu-se ao gráfico de dispersão para a apresentação dos resultados.

Dada a quantidade de informações e indicadores a serem analisadas, foram desenvolvidos três *dashboard* distintos, um para a classificação, um para a previsão e um outro para a segmentação de desempenho. Cada *dashboard* é composto pelos indicadores descritos anteriormente.

Após terem sido definidas as técnicas de visualização para a exibição de cada um dos indicadores, foi necessário organizá-las no *dashboard*. A posição dos indicadores tem impacto na percepção das informações que se exibem, os elementos no quadrante superior esquerdo e ao centro recebem a maior parte da atenção. Em seguida, os quadrantes superiores direito, inferior esquerdo e inferior direito, por esta ordem (Few, 2006b). Cada um dos *dashboards* foi modelado e implementado seguindo a seguinte lógica: o *dashboard* da classificação é composto por uma imagem e uma tabela em ambos os quadrantes superiores; o *dashboard* da previsão é composto por dois gráficos um com o desempenho geral do aluno e a previsão, que irá ocupar os quadrantes superiores, e outro gráfico constituído por vários subgráficos, com a previsão do desempenho por subdomínio, que ocupará os quadrantes inferiores, uma vez que é informação mais detalhada, com muita informação, pelo que não apresenta tanta relevância; por último, o *dashboard* de *clustering* que é composto por quatro gráficos, em que os gráficos com informação mais relevante do *dataset* atual, que permitem tirar conclusões, são apresentados nos quadrantes superiores e os restantes nos quadrantes inferiores. Na próxima secção apresentaremos estes *dashboards* de forma mais detalhada.

6.4 CARACTERIZAÇÃO E FUNCIONAMENTO DOS *dashboards*

Como já referido, foram desenvolvidos 3 *dashboards* diferentes, um para cada modelo implementado, nomeadamente: classificação, previsão e *clustering*. Pretendia-se, com cada um destes *dashboards*, fazer a monitorização do desempenho dos alunos ao longo do tempo. Esta análise pode ser feita por domínio, subdomínio ou aluno. O utilizador decide qual o período que pretende analisar bem como o tipo de análise que pretende fazer. Na Figura 27 podemos observar o dashboard que disponibiliza o resultado do modelo de classificação implementado. Como indicado anteriormente (Capítulo 5), para este processo de análise utilizámos o algoritmo RF. No *dashboard* correspondente (Figura 27) apenas se apresenta uma parte da árvore resultante de um dos estimadores. Para além disso, é apresentada uma tabela com alguma informação acerca do modelo implementado, tal como a *accuracy* obtida no treino e no teste, ou a precisão e a sensibilidade, bem como o número de estimadores e a profundidade máxima da árvore.

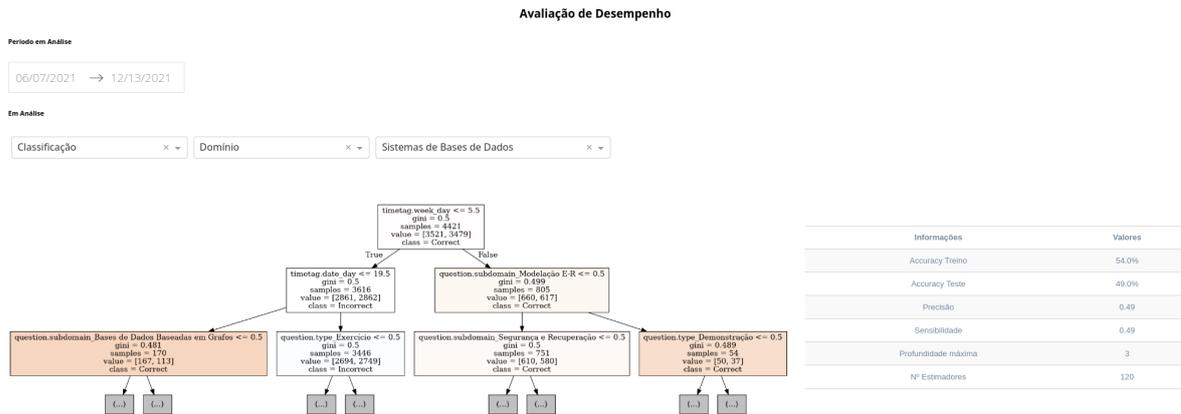


Figura 27: Dashboard de classificação

O dashboard apresentado na Figura 28 permite analisar o desempenho dos alunos, bem como perceber qual será o seu desempenho esperado num determinado período de tempo, de 5, 10 ou 15 dias. Além disso, o utilizador pode também definir qual o período de análise que quer utilizar. No primeiro gráfico, “Previsão do Desempenho Geral” é possível analisar a taxa de acerto, ou seja, o número de respostas certas sobre o número total de respostas. Desta forma, torna-se possível perceber qual a evolução dos alunos num determinado domínio ou subdomínio e a evolução de um determinado aluno. Para além disso, aquando da análise do desempenho por domínio e por aluno, é ainda possível observar o número de respostas que se prevê que este irá acertar por subdomínio, o que permitirá concluir quais as áreas que um determinado aluno (ou os alunos em geral) apresenta(m) maior dificuldade ou facilidade na sua aprendizagem.

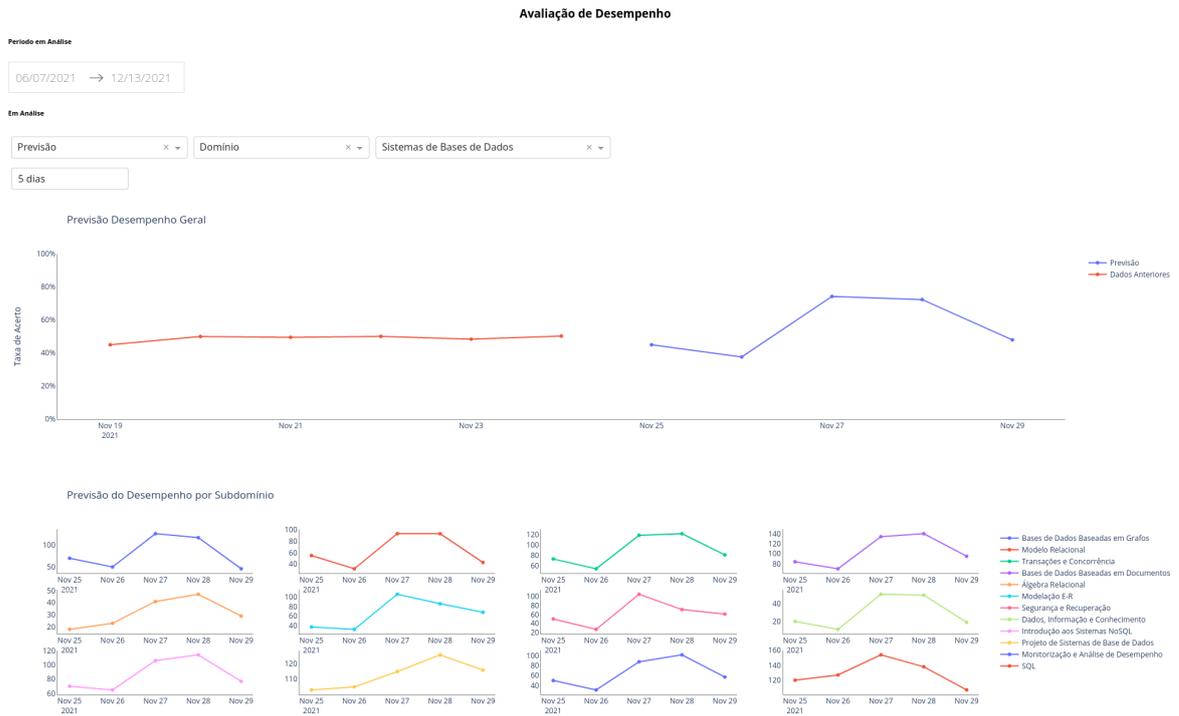


Figura 28: *Dashboard* de previsão

O último *dashboard* (Figura 29) apresenta os resultados obtidos do *clustering* do desempenho. Neste *dashboard* estão incluídos quatro gráficos, que permitem analisar os *clusters* obtidos e desta forma perceber as semelhanças e os padrões existentes nos dados. É possível analisar o agrupamento feito em relação ao tempo da resposta, dificuldade da questão e ainda relativamente ao subdomínio e tipo da questão.

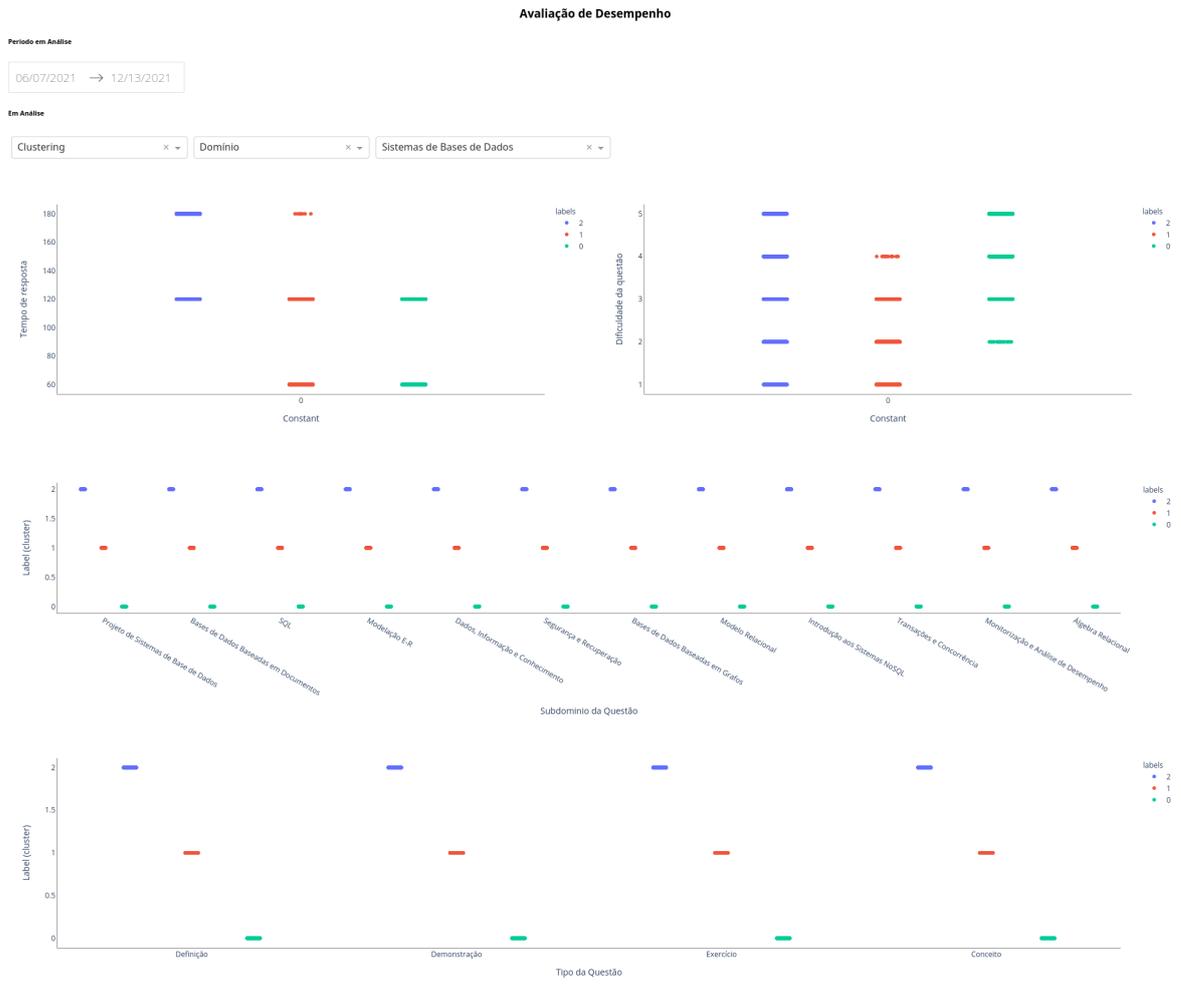


Figura 29: Dashboard de clustering

CONCLUSÕES E TRABALHO FUTURO

7.1 CONCLUSÕES

Atualmente, a avaliação do desempenho de alunos é considerada como uma parte integrante do processo de aprendizagem, uma vez que é algo que acontece ao longo da aprendizagem dos alunos. Além disso, professores e investigadores defendem que a capacidade de analisar e prever o desempenho do aluno é extremamente importante em ambientes educacionais. Através desta análise é possível revelar o conhecimento já adquirido pelos alunos e aquilo que ainda possam precisar de aprender ou de melhorar. Através da previsão do desempenho dos seus alunos ao longo do tempo, os professores podem identificar e analisar o processo de ensino e de aprendizagem e, desta forma, desenvolver uma estrutura de aprendizagem melhorada e mais adequada para cada um dos seus alunos. Por este motivo, várias instituições de ensino, tais como as universidades, procuram cada vez mais fornecer ferramentas de avaliação de desempenho que possam ser úteis tanto para os professores como para os seus alunos. Após o estudo de diversos sistemas de gestão de aprendizagem verificou-se que, na generalidade, estes não oferecem soluções de *dashboards* adequadas para as necessidades dos professores e alunos, uma vez que não permitem monitorizar em tempo real o desempenho dos alunos. Devido a isso, algumas instituições de ensino começaram a investir no desenvolvimento de soluções para a avaliação de desempenho dos seus alunos, suportadas por elementos de análise flexíveis, como painéis de visualização e de monitorização especialmente orientados para esses processos.

A aplicação de modelos de DM permite encontrar padrões, relações e anomalias em conjuntos de dados. Assim, é possível identificar, o mais cedo possível, eventuais questões problemáticas nos métodos e nos modelos de ensino ou em alunos que precisem de um acompanhamento especial e, conseqüentemente, atuar atempadamente na atenuação desse tipo de problemas. Uma possibilidade bastante concreta para o fazer é através de ações de investimento em processos de aprendizagem das instituições de ensino. Esta é uma solução considerada muito promissora para reduzir o fracasso e abandono escolar, existindo atualmente vários estudos nesta área pelo que já foram desenvolvidos vários modelos de avaliação de desempenho. Numa fase inicial deste trabalho de dissertação foram analisados modelos já desenvolvidos com o intuito de perceber quais as melhores metodologias e tecnologias para realizar o processo de construção do sistema de avaliação de desempenho pretendido.

Esta dissertação foi elaborada com base na metodologia CRISP-DM, desde a fase de conhecimento da área até à modelação e análise dos resultados. A primeira fase do trabalho envolveu a angariação e o pré-

processamento dos dados, tarefas de extrema importância em qualquer processo de análise de dados. Como tal, esta fase deve ser efetuada de forma ponderada, dadas as suas possíveis consequências, que tanto podem ser boas como más a longo prazo. Um correto pré-processamento permite não só ter um conhecimento absoluto do conjunto de dados como do problema em análise, uma vez que é responsável por efetuar todas as transformações no conjunto de dados, como, por exemplo, a remoção de dados não relevantes, ou a normalização dos dados, entre outros. Nesta fase foi importante fazer a seleção das *features* relevantes para o problema em questão, bem como o *encoding* dos atributos categóricos. Após o processamento dos dados foram explorados e implementados vários classificadores, isto porque para qualquer problema de aprendizagem automática é essencial utilizar mais do que um classificador – um classificador não é garantidamente o mais adequado para todos os problemas e dados. No nosso caso, os classificadores utilizados foram DT, RF e SVM. Com o intuito de otimizar os classificadores implementados recorreu-se à hiperparametrização, na qual foram aplicados os algoritmos *grid search* e *random search* com CV. Já no processo de previsão os dados foram classificados de acordo com a previsão de comportamentos futuros ou com valores estimados futuros, pelo que foi feita a previsão do desempenho dos alunos no futuro tendo em conta o histórico. O último algoritmo implementado foi o de *clustering*. A sua utilização permitiu encontrar agrupamentos naturais nos dados. Desta forma, através da criação de subconjuntos foi feita a separação dos dados potencialmente maliciosos dos dados normais. Das várias técnicas existentes, dada a natureza dos dados, aplicámos duas distintas, nomeadamente o *k-prototypes* e o *k-means*.

O sistema desenvolvido permite a monitorização do desempenho dos alunos em avaliações de questionários – *quizzes* –, que são compostos por vários domínios diferentes, que, por sua vez, são constituídos por inúmeros subdomínios. De modo a validar e analisar os modelos desenvolvidos foi utilizado um conjunto de dados constituído por dados sintetizados, gerados com um *script* que constituiu um conjunto de respostas bastante amplo, relativo a um período de 180 dias, tendo cada um dos dias cerca de 50 registos, contando com cerca de 340 000 entradas. Cada modelo desenvolvido foi testado e treinado com os dados de acordo com o pretendido, isto porque se pretendia uma análise bastante detalhada, em particular, por aluno, por domínio e por subdomínio de estudo. Todos os modelos base que foram desenvolvidos não apresentaram um bom desempenho. Tanto as DT como as RF obtiveram uma *accuracy* de treino de 60%/70% e de teste 50%. No caso das SVM a *accuracy*, tanto de treino como de teste, não passou dos 50%. A *accuracy* de treino variou um pouco dependendo do modelo implementado e das otimizações feitas. Contudo, os resultados obtidos não foram os esperados, uma vez que mesmo com otimização dos hiperparâmetros, com CV, não se verificou uma melhoria significativa do desempenho. Em alguns casos, piorou mesmo com uma descida da *accuracy* de treino em cerca de 10% não se verificando qualquer alteração na *accuracy* de teste. A natureza do conjunto de dados em análise, que foi gerado com métodos aleatórios, pode ter um grande impacto nos resultados obtidos uma vez que os dados não são de uma situação real. Desta forma, a qualidade destes dados, mesmo após o tratamento dos dados na fase do pré-processamento, terá comprometido os resultados dado que dificultou o treino e a generalização dos modelos. Este problema tornou-se evidente com os resultados obtidos no *clustering*, tanto com o método *k-prototypes* como com o *k-means*, uma vez que não foi possível encontrar uma relação direta entre as variáveis em estudo e identificar facilmente os *clusters*. Através da análise dos resultados obtidos foi perceptível a natu-

reza dos dados, que o conjunto de dados analisados eram sintetizados. Foi difícil tirar conclusões dos *clusters* obtidos, apenas tendo sido notória uma relação entre o tempo de resposta e a dificuldade da questão.

Após a modelação e implementação conceptual do sistema passou-se à implementação do sistema de *dashboards* adaptativos para apoiar e monitorizar o desempenho dos estudantes. Esta aplicação pretende auxiliar na tomada de decisões construtivas, informadas e conscientes, que permitam aos professores e alunos executar as tarefas de forma mais eficiente de modo a evitar e/ou reduzir o fracasso escolar. De modo permitir a monitorização dos alunos ao longo do tempo e auxiliar na análise e tomada de decisão recorremos aos *dashboards*. Para o desenvolvimento destes painéis de visualização utilizámos modelos, componentes e técnicas de visualização de dados que seguiram a literacia e padrões existentes. Foram implementados três *dashboards* para apresentar os resultados obtidos em fases anteriores, um para a classificação, previsão e clustering. O utilizador pode escolher o período que pretende analisar e o tipo de análise, que pode ser por domínio, subdomínio e aluno. Através destes painéis de visualização é possível observar padrões e sequências, sendo inferidas tendências no desempenho dos alunos. Assim, foram desenvolvidos *dashboards* que exibem os resultados obtidos pelos alunos e as previsões de desempenhos, mas também *dashboards* que permitem analisar o conjunto de dados em questão, através da observação e análise dos modelos de classificação e de *clustering*.

Por fim, refira-se que, nesta dissertação, pudemos verificar que as técnicas de *DM* em conjunto com os *dashboards* adaptativos ajudam a melhorar o ensino através da monitorização em tempo real do desempenho dos alunos. É, assim, possível analisar as perspetivas de sucesso e insucesso que um estudante tem, minimizando o risco de abandono escolar.

7.2 TRABALHO FUTURO

Os resultados obtidos pelos modelos implementados ficaram aquém das expectativas devido aos dados utilizados para teste e treino terem sido sintetizados, pelo que existem melhorias que podem levar à obtenção de resultados mais reais. Um dos próximos passos será integrar o sistema de avaliação de desempenho numa plataforma específica de avaliação de estudantes. Desta forma, seria possível analisar o comportamento do sistema desenvolvido com dados reais e verificar a sua *performance*. Será necessário também adaptar o design dos *dashboards* à plataforma de avaliação em questão.

Apesar de terem sido desenvolvidos vários modelos nesta dissertação, muitos outros podiam ser também concebidos e implementados. São várias as hipóteses de modelos a serem implementados futuramente. Contudo, um potencial candidato poderia ser as *Recurrent Neural Network (RNN)*, do tipo *Long Short-term Memory (LSTM)*, porque apresentam uma potencial de melhoria dos resultados dado que estas são utilizadas em modelos de previsão de séries cronológicas, permitindo prever valores futuros tendo por base dados anteriores sequenciais. Contudo, a utilização de *NN* não é garantidamente melhor do que os métodos de aprendizagem mais tradicionais, porque existem vários fatores que podem influenciar negativamente a capacidade de aprendizagem de modelos desse tipo como a dimensão e qualidade do conjunto de dados. O desenvolvimento destes modelos seria interessante para efeitos comparativos.

Nos *dashboards* atuais é permitido ao utilizador configurar algumas opções, como o tipo de *dashboard* que pretende analisar e filtrar os dados por domínio e subdomínio, bem como por aluno. Outra melhoria seria permitir uma maior configuração nos *dashboard*, como permitir a configuração do modelo desenvolvido por parte do utilizador, dando a possibilidade de alterar valores como os hiper parâmetros por exemplo.

Estas são algumas das possíveis linhas de trabalho que se poderiam adotar e desenvolver no futuro para melhorar a solução implementada.

BIBLIOGRAFIA

- License to evaluate: Preparing learning analytics dashboards for educational practice. 2018. ISBN 9781450364003. doi: 10.1145/3170358.3170421.
- A comparative study on decision tree and random forest using konstanz information miner (knife). *International Journal of Advanced Science and Technology*, 29:2365–2376, 4 2020. ISSN 22076360.
- Bassam Zafar Ahmed Mueen e Umar Manzoor. Modeling and predicting students' academic performance using data mining techniques. *International Journal of Emerging Trends in Engineering Research*, 8:36–42, 6 2016. doi: 10.30534/ijeter/2020/70862020.
- Abdulmohsen Algarni. Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7, 2016. ISSN 2158107X. doi: 10.14569/ijacsa.2016.070659.
- Sultan Saleh Ahmed Almekhlafy. Online learning of english language courses via blackboard at saudi universities in the era of covid-19: perception and use. *PSU Research Review*, 5, 2020. ISSN 2399-1747. doi: 10.1108/prr-08-2020-0026.
- Sylvain Arlot e Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 2010. ISSN 19357516. doi: 10.1214/09-SS054.
- ATutor. Atutor. <https://atutor.github.io/>, 2020. Accessed: 2021-12-03.
- John T. Avella, Mansureh Kebritchi, Sandra G. Nunn, e Therese Kanai. Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Journal of Asynchronous Learning Network*, 2016. ISSN 10928235. doi: 10.24059/olj.v20i2.790.
- James Bergstra e Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 2012. ISSN 15324435.
- Purnima Bholowalia e Arvind Kumar. Ebc-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105, 2014.
- Blackboard. Blackboard. <https://www.blackboard.com/>, 2018. Accessed: 2021-12-03.
- Marcel Ioan Bolos, Sabau-Popa Claudia Diana, e Bradea Ioana-Alexandra. Using dashboards in business analysis. *Annals of the University of Oradea: Economic Science*, 2014a. ISSN 1222-569X.
- Marcel Ioan Bolos, Sabau-Popa Claudia Diana, e Bradea Ioana-Alexandra. Using dashboards in business analysis. *Annals of the University of Oradea: Economic Science*, 2014b. ISSN 1222-569X.

- Nadire Cavus e Teyang Zabadi. A comparison of open source learning management systems. *Procedia - Social and Behavioral Sciences*, 2014. ISSN 18770428. doi: 10.1016/j.sbspro.2014.07.430.
- Claroline. Claroline. <https://www.claroline.com/>, 2020. Accessed: 2021-12-03.
- Corinna Cortes e Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20, 1995. ISSN 15730565. doi: 10.1023/A:1022627411411.
- Carolina Costa, Helena Alvelos, e Leonor Teixeira. The use of moodle e-learning platform: A study in a portuguese university. *Procedia Technology*, 5, 2012. ISSN 22120173. doi: 10.1016/j.protcy.2012.09.037.
- Bruno Carneiro da Rocha e Timoteo de Sousa Junior. Identifying bank frauds using crisp-dm and decision trees. *International Journal of Computer Science and Information Technology*, 2, 2010. ISSN 09754660. doi: 10.5121/ijcsit.2010.2512.
- Manoranjan Dash e Huan Liu. Feature selection for classification. *Intell. Data Anal.*, 1:131–156, 1997.
- Docebo. Docebo. <https://www.docebo.com//>, 2020. Accessed: 2021-12-03.
- Stephen Few. *Information Dashboard Design The Effective Visual Communication of Data*. 2006a. ISBN 0-596-10016-7.
- Stephen Few. Information dashboard design : Displaying data for at-a-glance monitoring. *Information dashboard design : displaying data for at-a-glance monitoring.*, 2006b.
- Abba Suganda Girsang, Fachrulhijriah Usman, e Rintismardika Sunarto. Clustering hostels data for customer preferences using k-prototype algorithm. *International Journal of Emerging Trends in Engineering Research*, 8:2650–2653, 6 2020. ISSN 23473983. doi: 10.30534/ijeter/2020/70862020.
- Chun houh Chen, Wolfgang Härdle, e Antony Unwin. *Handbook of Data Visualization*. 2008. doi: 10.1007/978-3-540-33037-0.
- Sasan Karamizadeh, Shahidan M. Abdullah, Mehran Halimi, Jafar Shayan, e Mohammad Javad Rajabi. Advantage and drawback of support vector machine functionality. 2014. doi: 10.1109/I4CT.2014.6914146.
- Mümine Kaya Keles e Selma Ayse Özel. A review of distance learning and learning management systems, 2016.
- G. Kesavaraj e S. Sukumaran. A study on classification techniques in data mining. 2013. doi: 10.1109/ICCCNT.2013.6726842.
- Brijesh Kumar e Saurabh Pal. Mining educational data to analyze students performance. *International Journal of Advanced Computer Science and Applications*, 2011. ISSN 2158107X. doi: 10.14569/ijacsa.2011.020609.
- Leonardo. Leonardo. <https://leonardo.di.uminho.pt/>, 2018. Accessed: 2021-11-30.

- Roberto Martínez Maldonado, Judy Kay, Kalina Yacef, e Beat Schwendimann. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. 2012. ISBN 9783642309496. doi: 10.1007/978-3-642-30950-2_62.
- M. Mayilvaganan e D. Kalpanadevi. Comparison of classification techniques for predicting the performance of students academic environment. volume 2015-March, 2015. doi: 10.1109/CNT.2014.7062736.
- Alaa M. Momani. Comparison between two learning management systems: Moodle and blackboard. *SSRN Electronic Journal*, 2012. ISSN 1556-5068. doi: 10.2139/ssrn.1608311.
- Moodle. Moodle. <https://moodle.org/>, 2020. Accessed: 2021-12-03.
- Mihaela Muntean, Gheorghe Sabau, Ana ramona Bologa, Traian Surcel, e Alexandra Florea. Performance dashboards for universities. 2010. ISBN 9789604742202.
- Carlos Márquez-Vera, Cristóbal Romero Morales, e Sebastián Ventura Soto. Predicting school failure and dropout by using data mining techniques. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 2013a. ISSN 19328540. doi: 10.1109/RITA.2013.2244695.
- Carlos Márquez-Vera, Cristóbal Romero Morales, e Sebastián Ventura Soto. Predicting school failure and dropout by using data mining techniques. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 2013b. ISSN 19328540. doi: 10.1109/RITA.2013.2244695.
- Emmanuel N. Ogor. Student academic performance monitoring and evaluation using data mining techniques. 2007. doi: 10.1109/CERMA.2007.4367712.
- Umesh Kumar Pandey e Saurabh Pal. Data mining : A prediction of performer or underperformer using classification. *CoRR*, abs/1104.4163, 2011. URL <http://arxiv.org/abs/1104.4163>.
- Yeonjeong Park e Il Hyun Jo. Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 2015. ISSN 09486968.
- Kedar Potdar, Taher S., e Chinmay D. A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computer Applications*, 175, 2017. doi: 10.5120/ijca2017915495.
- pypi. Kmodes. <https://pypi.org/project/kmodes/>, 2021. Accessed: 2022-01-30.
- Gabriel Reimers e Anna Neovesky. Student focused dashboards: An analysis of current student dashboards and what students really want. 2015. ISBN 9789897581076. doi: 10.5220/0005475103990404.
- N. Ruiz Reyes, P. Vera Candeas, S. García Galán, R. Viciania, F. Cañadas, e P. J. Reche. Comparing open-source e-learning platforms from adaptivity point of view. 2009. doi: 10.1109/EAAEIE.2009.5335482.
- Cristobal Romero e Sebastian Ventura. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3:12–27, 1 2013. ISSN 19424787. doi: 10.1002/widm.1075.

- Cristóbal Romero e Sebastian Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33:135–146, 07 2007. doi: 10.1016/j.eswa.2006.04.005.
- Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo, e César Hervás. Data mining algorithms to classify students. 2008a. ISBN 9780615306292.
- Cristóbal Romero, Sebastián Ventura, e Enrique García. Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, 2008b. ISSN 03601315. doi: 10.1016/j.compedu.2007.05.016.
- Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, e Danyel Fisher. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25, 2019. ISSN 19410506. doi: 10.1109/TVCG.2018.2864903.
- scikit learn. Decisiontreeclassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>, 2022a. Accessed: 2022-01-30.
- scikit learn. Kmeans. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>, 2022b. Accessed: 2022-01-30.
- scikit learn. Labelencoder. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>, 2022c. Accessed: 2022-01-30.
- scikit learn. Onehotencoder. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>, 2022d. Accessed: 2022-01-30.
- Amirah Mohamed Shahiri, Wahidah Husain, e Nur'Aini Abdul Rashid. A review on predicting student's performance using data mining techniques. 2015. doi: 10.1016/j.procs.2015.12.157.
- Ming Yi Shihm, Jar Wen Jheng, e Lien Fu Lai. A two-step method for clustering mixed categorical and numeric data. volume 13, 2010. doi: 10.6180/jase.2010.13.1.02.
- Shaymaa E. Sorour, Tsunenori Mine, Kazumasa Goda, e Sachio Hirokawa. A predictive model to evaluate student performance. *Journal of Information Processing*, 2015. ISSN 18826652. doi: 10.2197/ipsjip.23.192.
- Stephanie D. Teasley. Student facing dashboards: One size fits all? *Technology, Knowledge and Learning*, 2017. ISSN 22111670. doi: 10.1007/s10758-017-9314-3.
- Miller Tim. *Find the Best Visualizations for Your Metrics – How to Design a Dashboard*. 2019. doi: 10.1007/978-3-540-33037-0.
- Rüdiger Wirth. Crisp-dm : Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000.

Jia Wu, Xiu Yun Chen, Hao Zhang, Li Dong Xiong, Hang Lei, e Si Hao Deng. Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17, 2019. ISSN 1674862X. doi: 10.11989/JEST.1674-862X.80904120.

Rui Xu e Donald Wunsch li. Survey of clustering algorithms iee transactions on neural networks. *Ieee Transactions on Neural Networks*, 16:645–678, 2005.