



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sucesso escolar: abordagem de Business Analytics

Maria Francisca Maia Delfina

Mestrado em Métodos Analíticos para a Gestão (Business Analytics)

Orientadora:

PhD, Rute Gabriel Monteiro Saraiva, Professora Auxiliar Convidada,
ISCTE-IUL

Co-Orientadora:

PhD, Eva Patrícia Duarte Gonçalves, Professora Auxiliar Convidada,
ISCTE-IUL

Outubro, 2024



Departamento de Métodos Quantitativos para a Gestão e Economia

Sucesso escolar: abordagem de Business Analytics

Maria Francisca Maia Delfina

Mestrado em Métodos Analíticos para a Gestão (Business Analytics)

Orientadora:

PhD, Rute Gabriel Monteiro Saraiva, Professora Auxiliar Convidada,
ISCTE-IUL

Co-Orientadora:

PhD, Eva Patrícia Duarte Gonçalves, Professora Auxiliar Convidada,
ISCTE-IUL

Outubro, 2024

Dedicado a toda a minha família e amigos.

Agradecimentos

Gostaria de iniciar estes agradecimentos expressando o meu profundo obrigado a todos aqueles que, de uma forma ou de outra, estiveram ao meu lado ao longo desta jornada.

Aos meus pais, pela constante fonte de apoio, incentivo e amor incondicional. A vossa confiança em mim, ao longo de todas as etapas desta jornada, foi o alicerce que me manteve firme nos momentos mais desafiantes. Sou eternamente grata pela educação e valores que me transmitiram, que me permitem ser quem sou hoje.

Aos meus amigos, pela paciência e pelas palavras de incentivo ao longo de todo este percurso. A vossa amizade e presença foram essenciais para manter o equilíbrio e a motivação, mesmo nos momentos mais desafiantes.

Ao João, pelo amor, paciência, compreensão e presença incondicional. O teu apoio foi fundamental, especialmente nos momentos em que me senti mais sobrecarregada. Obrigada por acreditares em mim e por me dares a força necessária para continuar este caminho.

À professora Doutora Rute Saraiva, deixo um agradecimento muito especial. A sua orientação, paciência e compromisso foram essenciais para o desenvolvimento deste trabalho. O seu profissionalismo e dedicação, assim como as valiosas orientações e o incentivo constante, ajudaram-me a crescer tanto profissionalmente quanto pessoalmente.

À professora Doutora Eva Gonçalves, sou profundamente grata pelo acompanhamento cuidadoso e pelas contribuições importantes que enriquecem este trabalho. O seu olhar atento e as sugestões sempre pertinentes ajudaram a tornar esta dissertação mais sólida e completa.

Por fim, gostaria de expressar um agradecimento especial à DGEEC, pelo apoio prestado na recolha dos dados, pois só assim foi possível dispor do ficheiro de dados que sustentou esta dissertação.

Cada palavra desta dissertação reflete não apenas o meu esforço, mas também o apoio e carinho de cada um de vós. Este trabalho é também vosso, e a vocês, eu dedico o meu mais sincero obrigado.

Francisca

Resumo

A educação desempenha um papel central na formação da sociedade, sendo um indicador crucial para o desenvolvimento de um país. No contexto português, a temática do insucesso escolar e do abandono precoce tem sido amplamente discutida ao longo dos anos. Considerando as inúmeras investigações realizadas sobre o sucesso escolar em Portugal, é possível compreender que existem algumas lacunas, no que diz respeito à sua análise. Até ao momento, muitos estudos têm adotado uma abordagem que, embora relevante, pode ser complementada com perspetivas adicionais no que concerne aos modelos estatísticos aplicados. Embora existam pesquisas que abrangem o país e utilizam modelos estatísticos avançados, é importante destacar a necessidade de promover investigações que explorem diferentes metodologias estatísticas, capazes de oferecer novas perspetivas e enriquecer a análise desta temática.

Esta dissertação visa analisar o sucesso escolar no ensino básico do sistema público em Portugal, tendo por base a utilização dos *Self-Organizing Maps* (SOM) e os dados recolhidos na Direção Geral de Estatísticas da Educação e Ciência (DGEEC). Os dados utilizados para alcançar os objetivos definidos abrangem o período temporal entre 2019-2020 até 2021-2022 e foram desagregados por escola compilando diversos indicadores avaliativos do sucesso escolar.

Os *outputs* do SOM permitiram agrupar as várias escolas por região face ao nível de sucesso escolar, bem como permitiram perceber quais as variáveis que mais influenciam o sucesso escolar.

A presente dissertação registou algumas limitações. Porém, oferece contribuições relevantes para a análise do sucesso escolar, especialmente no que se refere à aplicação de modelos não supervisionados.

Palavras-chave: Sucesso escolar; SOM; ensino básico público; desempenho dos alunos; escolas de Portugal Continental.

Classificação JEL: C38, I21.

Abstract

Education plays a central role in the formation of society, being a crucial indicator for a country's development. In the Portuguese context, the issue of school failure and early school leaving has been widely discussed over the years. Considering the numerous investigations carried out on school success in Portugal, it is possible to understand that there are some gaps when it comes to analyzing it. So far, many studies have adopted an approach that, although relevant, can be complemented with additional perspectives regarding the statistical models applied. Although there is research that covers the country and uses advanced statistical models, it is important to highlight the need to promote investigations that explore different statistical methodologies, capable of offering new perspectives and enriching the analysis of this topic.

This research aims to analyse school success in basic education in the public system in Portugal, based on the use of Self-Organizing Maps (SOM) and data collected from the Direção Geral de Estatísticas da Educação e Ciência. The data used to achieve the defined objectives covers the period between 2019-2020 and 2021-2022 and was disaggregated by school, compiling various indicators for evaluating school success.

The SOM outputs made it possible to group the different schools by region in relation to their level of school success, as well as making it possible to understand which variables most influence school success.

This dissertation registered some limitations. However, it offers relevant contributions to the analysis of school success, especially with regard to the application of unsupervised models.

Key words: School success; SOM; public basic education; student performance; schools in mainland Portugal.

JEL Classification: C38, I21.

Índice

Índice	ix
Índice de Tabelas	xi
Glossário de siglas	xiii
Capítulo 1. Introdução.....	1
1.1 Enquadramento do tema	1
1.2 Problema e questão de investigação	7
1.3 Objetivos e contributos.....	8
1.4 Estrutura do documento.....	10
Capítulo 2. Revisão da Literatura.....	11
2.1 Sucesso escolar: conceitos.....	11
2.2 Limitações dos indicadores tradicionais de sucesso escolar em Portugal.....	18
2.3 Introdução ao SOM	21
2.4 Potencialidades do SOM na análise do sucesso escolar em Portugal	23
Capítulo 3. Metodologia de Investigação.....	25
3.1 Metodologia adotada	25
3.2 Metodologia CRISP-DM.....	26
3.2.1 Compreensão do negócio	26
3.2.2 Compreensão dos dados.....	27
3.2.3 Preparação dos dados	31
3.2.4 Modelação	34
3.2.5 Avaliação dos resultados	37
3.2.6 Implementação	38
Capítulo 4. Resultados e Discussão.....	39
4.1 Compreensão da complexidade do problema	39
4.2 Análise das possibilidades de agrupamento e seleção do agrupamento.....	41
4.3 Determinação do nível de desempenho.....	44
4.4 Posicionamento das escolas face ao sucesso escolar	45
4.5 Análise das questões de investigação.....	47
Capítulo 5. Conclusões.....	49
5.1 Recomendações da investigação	49
5.2 Limitações da investigação.....	49
5.3. Sugestões para investigações futuras.....	50
Referências Bibliográficas	51
Anexos.....	57

Índice de Tabelas

Tabela 1- Síntese das principais definições do conceito de sucesso escolar.....	14
Tabela 2- Grelha de avaliação resumida.....	29
Tabela 3- Tipos de variáveis presentes na grelha de avaliação.....	30
Tabela 4- Síntese das variáveis transformadas.....	33
Tabela 5- Processo de seleção da amostra.....	34
Tabela 6- Conjugação dos parâmetros do SOM.....	36
Tabela 7- Média e desvio-padrão das observações das escolas.....	40
Tabela 8- Síntese dos vários cenários e cálculo das métricas de medição da performance...	41

Índice de Figuras

Figura 1- Diagrama do modelo CRISP-DM.....	26
Figura 2- Código- Eliminação das variáveis em Python.....	32
Figura 3- Código- Transformação de variáveis categóricas em numéricas em Python.....	32
Figura 4- Código- Estandardização e normalização das variáveis selecionadas em Python	32
Figura 5- Código- Preenchimento de missing values em Python.....	33
Figura 6- Código- Aplicação do MiniSom em Python.....	35
Figura 7- Código- Aplicação das métricas de avaliação da performance em Python.....	38
Figura 8- Processo de implementacao dos resultados objdos atraves do SOM.....	39
Figura 9- Demonstração das matrizes de agrupamento em Python.....	43
Figura 10- Ilustração da matriz escolhida em Python.....	44
Figura 11- Ilustração da matriz escolhida em Python.....	47

Glossário de siglas

ASE	Ação Social Escolar
BMU	Best Matching Unit
CRISP-DM	Cross Industry Standard Process for Data Mining
DGEEC	Direção Geral de Estatísticas da Educação e Ciência
DP	Desvio-padrão
EE	Encarregados de Educação
KDD	Knowledge Discovery from Databases
OCDE	Organização para a Cooperação e Desenvolvimento Económico
PEI	Programa Educativo Individual
PMSE	Programa Mais Sucesso Escolar
QE	Quantization error
SEMMA	Sample, Explore, Modify, Model, Assess
SOM	Self-Organizing Maps
TE	Topographic error

Capítulo 1. Introdução

Os Agrupamentos de Escolas, que lidam diariamente com a diversidade de alunos e diversos ciclos de ensino básico (1.º, 2.º e 3.º ciclo), enfrentam desafios significativos na supervisão e melhoria da qualidade educativa. A necessidade de garantir uma educação de qualidade, inclusiva e adaptada às necessidades individuais de cada aluno torna-se imperativa, especialmente num contexto de exigência para transições entre ciclos com sucesso e para a diversificação curricular. A introdução que se segue auxilia na compreensão do conceito de sucesso escolar no contexto educativo do ensino básico e de que forma é possível avaliá-lo mediante técnicas assentes em modelos não supervisionados, utilizando os *Self-Organizing Maps* (SOM).

1.1 Enquadramento do tema

A educação desempenha um papel crucial na evolução social, económica e cultural da sociedade, sendo um indicador fundamental para o desenvolvimento de um país. No contexto português, a problemática do insucesso escolar e do abandono precoce tem sido objeto de discussão ao longo dos anos. A pesquisa de Martins (2017) realça que o insucesso escolar tornou-se um assunto de preocupação quando, no século XX, o direito à educação foi incluído na Constituição da República Portuguesa em 1976, o que fez do sucesso escolar um tema amplamente discutido pela política e pela sociedade civil, onde “todos os portugueses têm direito à educação e à cultura, nos termos da Constituição da República” (Decreto-Lei n.º 46/86, de 14 de Outubro). O capítulo seguinte tem como objetivo enquadrar o sistema de ensino em Portugal, explorando as dinâmicas, desafios e políticas educacionais que moldam o sistema de ensino no nível básico.

O trabalho de Álvares e Calado (2014) oferece uma análise abrangente sobre o insucesso e o abandono escolar precoce em Portugal. A sua obra destaca os programas de apoio implementados ao longo dos anos para combater estas questões persistentes. A compreensão desses programas é crucial para avaliar a efetividade de intervenções passadas e identificar lacunas a serem preenchidas.

No início do século XXI, o Ministério da Educação lança, em 2009, o Programa Mais Sucesso Escolar (PMSE), com o objetivo de prevenir o insucesso e o abandono escolar precoce no ensino básico. O PMSE prevê alterações ao nível do modelo escolar tradicional, tornando possíveis diferentes formas de agrupar os alunos e de gerir os tempos e espaços de instrução. Este programa marca o início de uma nova geração de políticas educativas, mais centradas no apoio das iniciativas de cada estabelecimento escolar, e abre caminho para a emergência de

outros projetos por parte das escolas (numa lógica *bottom up*), que se auto organizaram para fazer face aos seus problemas (Vieira, 2013).

O combate ao abandono e ao insucesso escolar no ensino básico é uma prioridade fundamental para a melhoria do sistema educativo e para a promoção de um futuro mais equitativo e sustentável em Portugal (Pedroso, 2022). O insucesso escolar, que se refere ao não cumprimento das metas académicas estabelecidas, e o abandono escolar precoce¹, que diz respeito à desistência precoce dos estudos antes da conclusão do ciclo de ensino obrigatório, isto é o 12.º ano, têm consequências negativas significativas para os indivíduos e para a sociedade, incluindo menores oportunidades de emprego, salários mais baixos e maior probabilidade de exclusão social (Martins, 2017).

Com vista a delinear um caminho de melhoria progressiva do indicador de sucesso escolar no ensino básico, i.e. um indicador central do sistema educativo, assegurando que todas as crianças e jovens concluam os doze anos de escolaridade obrigatória com sucesso, baseada numa extensa variedade de aprendizagens e recusando a dualização precoce, está previsto:

1. “Desenvolvimento de planos de acompanhamento e recuperação para alunos com dificuldades educativas de aprendizagem;
2. Implementação de programas de tutoria e apoio individualizado;
3. Garantia de igualdade de oportunidades e de condições de acesso e sucesso para todos os alunos, independentemente da sua origem socioeconómica, étnica, ou cultural;
4. Implementação de medidas de apoio específico para alunos com necessidades educativas especiais;
5. Adoção de processos de avaliação contínua das práticas educativas e dos resultados escolares;
6. Envolvimento dos pais e encarregados de educação no processo educativo” (Decreto-Lei n.º 176/12, de 02 de Agosto).

Adicionalmente, a pesquisa de Benavente (1990) enriquece a discussão ao abordar diferentes perspetivas e políticas relacionadas ao insucesso escolar no contexto português. As

¹ Abandono escolar precoce: população idade entre os 18 e os 24 anos que possui apenas o nível mínimo do ensino secundário ou um nível inferior e que não frequenta qualquer estabelecimento de ensino nem segue nenhuma formação;

Abandono escolar: saída do sistema de ensino antes da conclusão da escolaridade obrigatória, dentro dos limites etários previstos na lei.

suas análises conceituais proporcionam dados relevantes para compreender as raízes do problema e como as abordagens foram moldadas ao longo do tempo.

“O insucesso escolar revelado pelo alargamento do acesso à então designada escola primária de toda a população em idade escolar, fenómeno relativamente recente na Europa, tem sido explicado por diversas teorias. Em primeiro lugar, a *teoria dos «dotes»* (desde o final da segunda guerra mundial até ao final dos anos 60), baseada em explicações psicológicas individuais. O sucesso/insucesso é justificado pelas maiores ou menores capacidades dos alunos, pela sua inteligência, pelos seus «dotes» naturais” (Martins, 2010).

Benavente (1990) comenta que existem algumas fases a serem analisadas sobre o insucesso escolar, a primeira é o que chamamos de dote natural. A segunda é a característica sociocultural.

“Em segundo lugar, afirmou-se a teoria do «handicap» sociocultural (desde o final dos anos 60, início de 70), baseada em explicações de natureza sociológica. O sucesso/insucesso dos alunos é justificado pela sua pertença social, pela maior ou menor bagagem cultural de que dispõem à entrada na escola. O cruzamento entre origem social/resultados escolares revela a existência de mecanismos mais vastos na produção do sucesso/insucesso, que não pode ser atribuído apenas a causas psicológicas individuais” (Martins, 2010).

Na perspetiva de Nunes (2023), o insucesso escolar em Portugal está intimamente relacionado com as desigualdades sociais e com a estruturação territorial do país. O estudo revela que, embora o contexto socioeconómico seja um fator determinante no desempenho escolar dos alunos, as disparidades não se limitam a uma simples divisão entre classes. Em vez disso, o estudo destaca a forma como os fatores locais, nomeadamente a densidade populacional, a segregação entre escolas e a coesão social, também têm um papel significativo na perpetuação do insucesso escolar. Nos municípios com maiores desigualdades socioeconómicas, o insucesso escolar tende a ser mais prevalente, devido, em parte, à segregação escolar entre diferentes contextos socioeconómicos. Isso resulta num ambiente educacional menos favorável para os alunos provenientes de famílias de menores rendimentos, que enfrentam uma série de desafios, como o menor acesso a recursos educativos e apoio familiar. Por outro lado, em municípios onde existe maior coesão social e melhores condições de emprego, os alunos de classes socioeconómicas mais baixas apresentam, em média, melhores resultados escolares. A presença de um setor secundário ativo e a estabilidade das estruturas familiares também contribuem para a melhoria do desempenho escolar, mesmo em contextos mais desfavorecidos. Dessa forma, Nunes e os seus colaboradores sugerem que o insucesso escolar não pode ser abordado apenas com políticas educativas gerais. Pelo contrário, é necessário um enfoque mais localizado, que tenha em consideração as especificidades

socioeconómicas e territoriais dos municípios, promovendo a equidade e a inclusão através de uma melhor distribuição de recursos e apoio especializado. O estudo defende que a diminuição do insucesso escolar passa por uma abordagem integrada, onde a educação, a economia e as políticas sociais trabalham juntas para melhorar as condições de vida e de aprendizagem dos jovens.

A obra coordenada por Renato Carmo (2024) oferece uma análise detalhada das desigualdades sociais que influenciam o insucesso escolar e as dificuldades de inserção no mercado de trabalho dos jovens em Portugal. Os autores destacam que as condições socioeconómicas determinam fortemente o desempenho escolar, fazendo com que muitos jovens abandonem precocemente a escola e posteriormente apresentem dificuldades no acesso a empregos estáveis e bem remunerados. A pesquisa sugere que as políticas públicas existentes não têm sido suficientemente eficazes para reduzir essas disparidades, especialmente no que diz respeito ao apoio direcionado aos jovens em risco. Posto isto, os autores propõem uma revisão das abordagens educacionais e de proteção social, de modo a garantir que todos os jovens tenham acesso a oportunidades igualitárias no sistema educativo e no mercado de trabalho. Este estudo permite avaliar a eficácia das políticas passadas e identificar lacunas que precisam de ser abordadas para um futuro mais inclusivo e justo para os jovens portugueses.

A escola é um “promotor determinante do desenvolvimento cónico e social do Ser Humano” (Miguel, Rijo, & Lima, 2012, p. 128), porém “políticas educativas, formação de professores, modelos pedagógicos, análises curriculares, dificuldades de aprendizagem, desenvolvimento cognitivo” (Benavente, 1990, p. 715), são fatores que podem influenciar os resultados dos alunos, tanto de uma maneira positiva, como negativa, sendo o exemplo do insucesso escolar (Martins, 2017). Segundo Martins (1993) a escola “democrática” com alunos heterogéneos tanto social, como culturalmente é um lugar onde todos são iguais e onde se introduz “currículos universais, conotados com um perfil médio de aluno e privilegiando um saber mais clássico, geral e enciclopédico, isto é, propõe objetivos pouco pragmáticos, tendo em conta as realidades diferenciadas que são os alunos que a frequentam” (Martins & Cabrita, 1993, citado por Martins, 2017). Nas discussões entre docentes sobre as possíveis causas do insucesso escolar, esse fenómeno “é visto como um problema do aluno, jamais, ou muito raramente como um fenómeno que possa envolver a organização e as práticas escolares” (Alves, 2010, citado por Martins, 2017). No entanto, o insucesso do aluno acaba por refletir, de certa forma, o insucesso da escola, pois as aprendizagens não são feitas, ou quando adquiridas, são insuficientes para permitir que o aluno consiga progredir no seu percurso escolar. Assim sendo, é “evidente que as responsabilidades devem ser partilhadas” (Martins, 2006, citado por Martins,

2017) visto que o insucesso escolar é “um fenómeno social, das realidades incontestáveis, divulgando nas escolas, famílias, a nível político e nos meios de comunicação social” (Rosa, 2013, citado por Martins, 2017). Desta forma, a escola, enquanto promotora de aprendizagens, deve considerar a singularidade de cada criança e ajustar os seus currículos de acordo com essa diversidade.

Tendo em conta que o insucesso escolar reflete tanto fatores externos como internos às práticas pedagógicas, torna-se fundamental analisar a gestão escolar como um elemento estruturante que pode amplificar ou mitigar tais desafios, especialmente no contexto do sistema educativo português.

Ventura, Castanheira e Adelino Costa (2006) exploram assim a gestão das escolas em Portugal, trazendo à tona elementos cruciais que influenciam o ambiente educacional. A compreensão da gestão escolar é essencial para identificar fatores internos que podem afetar o sucesso escolar. Analisar como as escolas são geridas pode fornecer informações sobre práticas eficazes e desafios a serem superados. “No quadro do sistema educativo português tem sido insistentemente brandida em termos discursivos a bandeira da descentralização, da autonomia, da territorialização das políticas educativas e da gestão escolar” (Ventura et al, 2006).

Nesse sentido, a gestão escolar em Portugal tem evoluído através de reformas significativas, que visam concretizar os princípios de descentralização e autonomia, que se reflete diretamente no modelo de organização e liderança das escolas.

Em Portugal, o modelo de gestão escolar sofreu uma transformação significativa com a publicação do Decreto-Lei n.º 75/2018, que consolidou o papel do Diretor como órgão unipessoal responsável pela totalidade da gestão e funcionamento da escola. Este cargo passou a ser regulado pelo Conselho Geral, que detém competências cruciais como a eleição, avaliação, destituição do Diretor e a aprovação dos documentos estruturantes por ele elaborados. Este decreto representa um marco no processo de reforço da autonomia das escolas, sendo uma das principais vertentes da descentralização educativa no país, juntamente com a transferência de competências para os territórios. O objetivo central dessa reforma prende-se com o aprimoramento da qualidade do serviço educativo, através da proximidade entre as escolas e as comunidades locais. Com uma gestão mais ágil e adaptada às especificidades dos contextos regionais, torna-se possível responder de forma mais eficaz aos desafios, como o combate ao insucesso escolar e ao abandono escolar precoce, fortalecendo as parcerias entre escolas e comunidades.

“Ao director é confiada a gestão administrativa, financeira e pedagógica, assumindo, para o efeito, a presidência do conselho pedagógico. Exercendo também competências no domínio

da gestão pedagógica, sem as quais estaria sempre diminuído nas suas funções, entende-se que o director deve ser recrutado de entre docentes do ensino público ou particular e cooperativo qualificados para o exercício das funções, seja pela formação ou pela experiência na administração e gestão escolar” (Decreto-Lei n.º 75/08, de 22 de Abril).

Como parte dos avanços na gestão escolar, foram também implementadas reformas nas carreiras dos docentes, com o objetivo de fortalecer a organização e a qualidade do ensino público. Essas medidas têm como finalidade dotar as escolas de um corpo docente mais experiente e qualificado, capaz de assumir funções de maior responsabilidade e contribuir para um gestão escolar mais eficiente e alinhada à missão do serviço público da educação.

“O Governo promoveu ainda a alteração do Estatuto da Carreira dos Educadores de Infância e dos Professores dos Ensinos Básico e Secundário, no sentido de dotar cada estabelecimento de ensino público de um corpo de docentes reconhecido, com mais experiência, mais autoridade e mais formação, que assegure em permanência funções de maior responsabilidade. A estruturação da carreira, com a criação da categoria de professor titular, à qual são reservadas as actividades de coordenação e supervisão, constituiu um importante contributo para a capacidade de organização das escolas em função da missão de serviço público que lhes está confiada” (Decreto-Lei n.º 75/08, de 22 de Abril).

Lemos (2014) apresenta uma visão aprofundada sobre a influência da Organização para a Cooperação e Desenvolvimento Económico (OCDE) nas políticas públicas de educação em Portugal. A contextualização internacional é fundamental para entender como as diretrizes da OCDE moldaram o sistema educacional português. Essa perspetiva internacional permite uma análise mais abrangente dos indicadores de sucesso escolar.

Os diversos estudos realizados ao longo dos anos pela OCDE e pela União Europeia têm em comum, para além da contemporaneidade, o facto de considerarem a equidade como uma questão central das políticas de educação, olhando-a, não só como um objetivo, mas também como um instrumento essencial das políticas públicas. Ambos consideram a equidade em educação como um fim em si mesmo, porque o direito ao desenvolvimento das capacidades de cada pessoa e à sua participação na sociedade faz parte dos direitos inscritos na maioria das leis fundamentais dos diferentes países, e ambos consideram também que a equidade em educação é uma condição indispensável à consecução de outros fins, quer sociais quer económicos, e, portanto, à formulação das políticas públicas (Lemos, 2014, p. 152).

É essencial consolidar a trajetória de redução do insucesso escolar no ensino básico e continuar a avançar num caminho de melhoria progressiva deste indicador central do sistema educativo, assegurando que todas as crianças e jovens concluam os primeiros anos de

escolaridade com uma educação de qualidade, baseada numa ampla variedade de aprendizagens e rejeitando a dualização precoce (Lei n.º 7B/16, de 31 de Março). No que diz respeito aos métodos de análise de dados, a avaliação da qualidade do ensino tem, de maneira geral, utilizado estatísticas descritivas, nomeadamente medidas de localização (e.g., quartis), o que dificulta uma compreensão mais aprofundada dos dados e uma identificação mais precisa dos diferentes tipos de casos que existem. Este *gap* na análise do sucesso escolar em Portugal evidencia a importância de explorar e adotar metodologias alternativas. Embora existam estudos que recorrem sobretudo a análises fatoriais, os quais têm sido bastante eficazes na identificação das causas do insucesso escolar, o ponto-chave deste estudo reside na adoção de uma metodologia diferente, que pelas suas características, possibilita outro tipo de análise e de resultados.

1.2 Problema e questão de investigação

A investigação sobre o sucesso escolar no ensino básico do sistema público em Portugal é um campo complexo e multifacetado. Neste contexto, uma questão de investigação relevante é: como a aplicação do SOM pode contribuir para a análise do sucesso escolar?

Para responder a esta questão, é crucial entender os desafios e nuances envolvidos na análise do sucesso escolar. Diversos estudos, como os de Afonso (2009) e Almada (2015), destacam que o sucesso escolar não pode ser totalmente capturado por medidas padronizadas e rankings, devido à sua natureza multifacetada e contextual.

As desigualdades socioeconómicas, discutidas por Álvares e Calado (2014) e Alves e Cabral (2016a), representam um dos principais obstáculos para o sucesso escolar. A falta de recursos em certas áreas, especialmente nas regiões rurais, e as disparidades no acesso à educação entre diferentes grupos socioeconómicos, contribuem para um cenário desigual.

Outro fator que deve ser considerado é o impacto da imigração, que está a movimentar cada vez mais crianças e jovens com outros idiomas e culturas para o sistema educativo português. Este fenómeno acrescenta uma camada adicional de complexidade ao processo de promoção do sucesso escolar, pois essas crianças e jovens frequentemente enfrentam barreiras linguísticas e culturais que podem dificultar a sua integração e desempenho académico. Estudos como os de Pereira (2021) e Madeira (2023) destacam a necessidade de adaptação dos métodos de ensino e de suporte adicional para esses alunos, de modo a garantir que todos tenham a oportunidade de alcançar o sucesso escolar.

Além disso, questões culturais e de género, conforme discutido pelo Conselho Nacional de Educação (2017) e Ribeiro et al. (2006), desempenham um papel significativo. Estereótipos de

género podem influenciar as expectativas e o tratamento dos alunos, afetando o seu desempenho e motivação.

A análise tradicional desses problemas, normalmente baseada em estatística descritiva, enfrenta muitas vezes limitações na identificação de padrões complexos nos dados. É aqui que a aplicação do SOM pode oferecer uma perspetiva inovadora. O SOM, como discutido por Negri (2013), é uma técnica poderosa de análise de dados que pode identificar padrões não triviais em conjuntos de dados complexos.

Ao aplicar o SOM na análise do sucesso escolar, podemos identificar padrões e correlações entre diferentes variáveis, como desempenho académico, contexto socioeconómico e características individuais dos alunos. Isso pode fornecer dados valiosos para a formulação de políticas educacionais mais eficazes e medidas de intervenção direcionadas.

No entanto, é importante reconhecer que a aplicação do SOM também apresenta desafios e limitações. Uma avaliação crítica dos resultados obtidos é essencial, conforme destacado por Roazzi e Almeida (1988), para garantir a validade e a aplicabilidade das conclusões.

Ao considerar os objetivos de investigação, é imperativo avaliar a efetividade do SOM como ferramenta de análise do sucesso escolar. A literatura oferece uma base sólida para entender o contexto educacional português, para além de permitir identificar os principais indicadores que afetam o sucesso escolar, considerando as especificidades do nível básico. Com este conhecimento é possível realizar uma avaliação mais informada da utilidade e aplicabilidade do SOM neste cenário. Além disso, a análise das zonas geográficas com maior sucesso escolar pode fornecer dados sobre fatores regionais que influenciam o desempenho dos alunos.

Outro ponto crucial a ser explorado é a avaliação da efetividade do SOM como ferramenta de análise do sucesso escolar. Esta avaliação é vital pois permite compreender de que forma se pode aplicar o SOM à análise do sucesso escolar, bem como as suas vantagens e desvantagens.

1.3 Objetivos e contributos

O sucesso escolar é um tema de grande relevância no contexto educacional contemporâneo, refletindo não apenas o desempenho académico dos alunos, mas também uma série de dinâmicas sociais, económicas e culturais que influenciam o processo de aprendizagem. No âmbito do ensino básico em Portugal, compreender as causas da diversidade do sucesso escolar é fundamental para o desenvolvimento de políticas educacionais mais eficazes e inclusivas.

Neste contexto, esta pesquisa propõe-se a investigar detalhadamente as variáveis que contribuem para o sucesso escolar no ensino básico e a aplicar técnicas de análise baseadas em aprendizagem de máquina, nomeadamente o SOM, que permitem compreender o agrupamento das mesmas. Como objetivos específicos a dissertação visa compreender as similitudes que determinam o agrupamento do sucesso escolar e avaliar a aplicação do SOM no estabelecimento de ferramentas escaláveis com a atualização permanente de variáveis.

Como objetivo geral definiu-se: avaliar as causas da diversidade ou as similitudes do sucesso escolar no ensino básico, uma vez que o SOM é uma técnica de classificação e agrupamento.

Assim, os objetivos concretos que orientam esta dissertação são:

1. Analisar e identificar as principais variáveis e indicadores que contribuem para o sucesso escolar no nível básico de ensino. Isso inclui fatores socioeconómicos, características individuais dos alunos e contextos educacionais.
2. Investigar e mapear as áreas geográficas onde se regista um maior sucesso escolar. Isso permitirá uma compreensão mais aprofundada das disparidades regionais e ajudará na formulação de políticas educacionais mais direcionadas.
3. Avaliar a eficiência do SOM como ferramenta de análise do sucesso escolar. Isso envolve explorar como o SOM pode ser aplicado na análise do sucesso escolar e identificar as suas vantagens e desvantagens.

Acredita-se que os resultados desta dissertação não só contribuirão para o avanço do conhecimento académico sobre o sucesso escolar, mas também terão implicações práticas significativas para a formulação de políticas educacionais e práticas pedagógicas. Ao abordar questões complexas como desigualdades socioeconómicas e a eficácia de ferramentas analíticas avançadas, este estudo estará na vanguarda da pesquisa educacional contemporânea. Assim:

1. Ao identificar as principais variáveis e indicadores que influenciam o sucesso escolar, esta dissertação contribuirá para uma compreensão mais profunda dos mecanismos subjacentes ao desempenho dos alunos. Isso pode informar políticas educacionais mais eficazes e intervenções direcionadas.
2. O mapeamento das zonas de sucesso escolar fornecerá informações valiosas para os decisores políticos e as autoridades educacionais. Isso ajudará na alocação de recursos e no desenvolvimento de estratégias específicas para regiões com desafios educacionais distintos.
3. Ao avaliar a eficiência do SOM na análise do sucesso escolar, esta dissertação contribuirá para a validação e refinamento contínuo dessa técnica analítica. Isso pode

abrir novas perspectivas para a aplicação de métodos de análise de dados avançados na educação.

1.4 Estrutura do documento

A presente dissertação está estruturada em cinco capítulos que refletem os distintos estágios de investigação. No primeiro capítulo são apresentados o tema, as lacunas e as questões de investigação identificadas, além dos objetivos e das principais contribuições que esta investigação oferece à comunidade científica. O segundo capítulo, denominado por revisão da literatura, apresenta o enquadramento teórico, onde a investigação recai sobre o conceito de sucesso escolar, os principais indicadores e variáveis que o afetam e, por fim, a forma como o sucesso escolar tem sido avaliado ao longo dos anos. O terceiro capítulo apresenta a metodologia adotada e as diversas etapas que a constituem, desde a compreensão do negócio até à fase de implementação. No capítulo seguinte, o quarto, são expostos os resultados obtidos com a aplicação do modelo, acompanhados de uma análise e discussão sobre os mesmos. Por fim, o quinto capítulo expõe as conclusões da dissertação.

Capítulo 2. Revisão da Literatura

Este capítulo aborda aspetos teóricos relacionados com a forma como se interpreta o conceito de sucesso escolar, quais os principais indicadores que o afetam e, ainda, como se pode analisar o sucesso escolar ao nível do ensino básico. Mais especificamente, são abordados os modelos de avaliação mais comuns e quais as técnicas de avaliação. Neste sentido, o principal objetivo é contribuir para a identificação de oportunidades de melhoria na análise do sucesso escolar.

2.1 Sucesso escolar: conceitos

O sucesso escolar é um indicador crítico da eficácia do sistema educacional, refletindo não apenas no desempenho do aluno e nas práticas dos docentes, mas também nas dinâmicas organizacionais e políticas que permeiam o contexto educacional. Analisando os desafios e indicadores relacionados ao sucesso escolar em Portugal, observamos uma variedade de perspetivas e abordagens presentes na literatura.

Ao longo das décadas, o conceito de sucesso escolar tem evoluído e sido interpretado de várias maneiras por académicos e educadores. Desde os estudos de Coleman et al. (1966), que relacionaram o sucesso escolar com o ambiente familiar e socioeconómico do aluno, até às propostas mais recentes de Ventura (2021), que enfatizam a importância de criar um ambiente educativo inclusivo e equitativo, tem havido uma diversidade de abordagens.

No início da década de 1970, Figueiredo (1975) propôs que o sucesso escolar estivesse associado ao alcance de um desempenho escolar satisfatório em todas as disciplinas, enquanto Rutter et al. (1979) consideraram-no como a capacidade de atingir um nível adequado de competências escolares e adaptação psicossocial. Epstein (1983), por sua vez, destacou a importância de uma parceria entre a escola, a família e a comunidade para apoiar o desenvolvimento do aluno. A obra de Roazzi e Almeida (1988) questiona se o insucesso escolar é resultado do aluno ou do sistema escolar. Esta abordagem crítica é essencial para compreender as diferentes perspetivas que podem moldar as discussões sobre o sucesso escolar. A reflexão sobre a responsabilidade do aluno versus a responsabilidade do sistema é crucial para desenvolver estratégias eficazes de intervenção.

Na década de 1990 encontra-se um vasto conjunto de estudos sobre a temática, incidindo em focos de análise distintos. Muitos destes consideraram as variáveis pessoais dos estudantes, sejam elas as suas aprendizagens prévias, a saúde, as competências sociais, as atitudes e crenças face ao seu desempenho, as habilidades cognitivas e as dimensões motivacionais (Benavente, Cortesão & Torres, 1990; Fernandes, 1991; Névoa, 1992; Murnane & Levy, 1996; Peixoto,

1999, citado por Soares, 2015). Outros estudos incidem nas variáveis associadas à escola e ao ensino, como o clima da escola, as características do professor, o currículo, as práticas pedagógicas, a estrutura de sala de aula, os sistemas de apoio e os recursos disponibilizados (Eccles et al., 1993; Rangel, 1994, citado por Soares, 2015).

Avançando para os anos 2000, investigadores como Alarcão (2001), Guthrie et al. (2004) e Canário (2005) consideraram que o sucesso escolar apresenta-se intrinsecamente relacionado com a capacidade do sistema educativo proporcionar oportunidades iguais a todos os alunos, de forma que estes atinjam os seus objetivos pessoais e académicos. Jimerson et al. (2000) e Lareau (2003) discutiram que o sucesso escolar é visto como a capacidade dos alunos navegarem nos sistemas educacionais, especialmente em contextos de desigualdade social. No que diz respeito a Crosnoe e Johnson (2011), estes autores argumentaram que o sucesso escolar é influenciado por uma interação complexa de fatores sociais, familiares e académicos. No que concerne a Justino et al. (2014), estes afirmaram que o sucesso escolar é alcançado quando os alunos demonstram competências académicas e habilidades socioemocionais. Por outro lado, Verdasca (2016) afirmou que o sucesso escolar é medido pelo grau de alcance dos objetivos educativos estabelecidos, considerando tanto os resultados académicos como o desenvolvimento integral do aluno.

De facto, as definições de sucesso escolar têm evoluído ao longo do tempo, refletindo uma compreensão cada vez mais holística e contextualizada desse conceito fundamental na educação.

Mais recentemente, o trabalho de Amorim (2019), por exemplo, proporciona uma visão aprofundada sobre os fatores organizacionais e a sua relação com o (in)sucesso escolar. Através de um estudo de caso duplo comparativo, a autora explora como as características organizacionais das instituições de ensino podem influenciar diretamente o desempenho dos alunos. Esta análise contribui significativamente para a compreensão de como as estruturas escolares impactam os resultados educacionais. Amorim (2019) cita os fatores ditos por Cabral (2014) e coloca-os num quadro para reconhecimento (Quadro 1).

Quadro 1: Fatores organizacionais que influenciam o sucesso escolar (Cabral, 2014, citado por Amorim, 2019)

Categorias	Subcategorias
1. Liderança administrativa e organizacional	1.1. liderança eficaz do pessoal, pautada pela firmeza e determinação 1.2. Abordagem participativa, convocando-se os docentes a participar ativamente no estabelecimento de normas e finalidades claras para cada escola 1.3. Promoção de estratégias concertadas de atuação em projetos de trabalho 1.4. Liderança instrucional ou pedagógica
2. Clima de escola/culturas profissionais	2.1. Expectativas elevadas face ao desempenho escolar dos alunos 2.2. Envolvimento dos alunos nas metas e objetivos a atingir 2.3. Responsabilização dos alunos pelos seu próprio processo de aprendizagem 2.4. Comunicação entre os vários intervenientes no processo educativo (direção, docentes, alunos, pais e comunidade) 2.5. Trabalho colaborativo e relações de colegialidade entre professores 2.6. Estabilidade do pessoal docente 2.7. Formação de professores articulada com o projeto educativo de cada escola 2.8. Ligação da escola à comunidade 2.9. Abertura da escola à comunidade
3. Organização e gestão de tempos, espaços e recursos	3.1. Organização dos tempos e espaços escolares 3.2. Regime anual de reprovação/aprovação 3.3. Distribuição dos professores por turma
4. Organização ao nível dos alunos	4.1. Critérios de distribuição de alunos por turma 4.2. Dimensão e caracterização das turmas
5. Práticas pedagógicas/estratégias de ensino	5.1. Currículo essencial e viável 5.2. Clima de sala de aula estruturado e securizante 5.3. Monitorização sistemática do progresso dos alunos na aprendizagem 5.4. Reforço positivo dos alunos face às aprendizagens realizadas 5.5. Relação professor-aluno

A pesquisa de Martins (2017) aborda a prevenção e intervenção no insucesso escolar, focando na educação pré-escolar e no 1º ciclo do ensino básico. A dissertação oferece visões valiosas sobre estratégias que podem ser implementadas no início da jornada educacional para evitar possíveis obstáculos no percurso do aluno. A prevenção do insucesso escolar é um elemento crucial na busca por um sistema educacional mais eficiente. Citando não só o papel do educador nessas estratégias, mas como práticas de desporto e uma vida mais saudável também, física e mental.

O insucesso escolar representa um problema que impacta toda a sociedade e cujos números têm aumentado. De forma a lidar com esta questão, o Ministério da Educação implementou medidas previstas no Decreto-Lei n.º 176/2012, que visam prevenir o insucesso escolar. De acordo com este diploma, quando um aluno enfrenta dificuldades de aprendizagem, devem ser aplicadas ações específicas.

“Através do reforço das medidas de apoio ao estudo, que garantam um acompanhamento mais eficaz do aluno face às primeiras dificuldades detetadas” e “de um acompanhamento extraordinário dos alunos estabelecido no calendário escolar.” Adoção de “percursos curriculares alternativos e programas integrados de educação e formação, adaptadas ao perfil e especificidades dos alunos”, entre outras medidas (Martins, 2017; pp. 14).

Mendonça (2006) contribui para a discussão ao explorar a problemática do insucesso escolar na Madeira, analisando a escolaridade obrigatória no arquipélago em finais do século XX. Esta perspetiva regional oferece uma visão única dos desafios específicos enfrentados em determinadas áreas geográficas, permitindo uma compreensão mais profunda das dinâmicas que afetam o sucesso escolar.

O progresso da escolarização, associado ao movimento revolucionário dos anos de 1974/75, foi além do simples objetivo de oferecer oportunidades de acesso iguais a um património cultural universal.

O acréscimo de escolarização verificado neste período, ao desenvolver uma maior implicação social da educação, induziu dinâmicas educativas em populações não escolarizadas (Mendonça, 2006).

Ao integrar as contribuições da literatura anterior, algumas das que abordaram o sucesso escolar no contexto português, este capítulo pretende construir uma base sólida para a compreensão dos desafios e indicadores relacionados ao sucesso educacional. A análise combinada desses estudos oferece uma visão abrangente das complexidades inerentes ao sistema educacional em Portugal.

As principais definições utilizadas para definir o conceito de sucesso escolar, pelos diversos autores citados ao longo do subcapítulo 2.1, estão sintetizadas na Tabela 1.

Tabela 1- Síntese das principais definições do conceito de sucesso escolar

Autor	Ano	Definição
Coleman et al.	1966	O sucesso escolar está relacionado com o ambiente familiar e socioeconómico do aluno
Figueiredo	1975	O sucesso escolar é alcançado quando os alunos atingem um desempenho académico

		satisfatório em todas as disciplinas
Rutter et al.	1979	Considera o sucesso escolar como a capacidade de atingir um nível satisfatório de habilidades académicas e adaptação psicossocial
Epstein	1983	O sucesso escolar envolve uma parceria entre a escola, a família e a comunidade para apoiar o desenvolvimento do aluno
Roazi & Almeida	1988	O sucesso escolar é alcançado quando os alunos estão envolvidos nas atividades educacionais e têm um bom relacionamento com os professores e colegas
Benavente	1990	Define o sucesso escolar como o alcance de objetivos educacionais e desenvolvimento integral do aluno, incluindo aspectos sociais e emocionais
Cortesão & Torres	1990	Definem o sucesso escolar como o alcance de um nível satisfatório de desempenho académico e adaptação psicossocial na escola
Fernandes	1991	Considera o sucesso escolar como o alcance de habilidades académicas e adaptação social
Névoa	1992	Considera o sucesso escolar como sendo a aquisição de conhecimentos, competências, e também o desenvolvimento pessoal e social do aluno

Eccles et al.	1993	Considera o sucesso escolar como a realização acadêmica do aluno, incluindo notas, testes padronizados e conclusão do ensino médio
Rangel	1994	Define sucesso escolar como a conclusão dos estudos com êxito, adquirindo conhecimentos e competências relevantes
Murnane & Levy	1996	Enfatizam a importância das habilidades acadêmicas e cognitivas para o sucesso escolar e a empregabilidade futura dos alunos
Peixoto	1999	Considera o sucesso escolar como a capacidade do aluno de atingir metas educacionais pessoais e o desenvolvimento de habilidades sociais
Jimerson et al.	2000	O sucesso escolar é visto como o alcance de competências acadêmicas e comportamento positivo na escola
Alarcão	2001	Considera o sucesso escolar como a capacidade do sistema educativo proporcionar oportunidades iguais a todos os alunos
Lareau	2003	Considera o sucesso escolar como a capacidade dos alunos navegarem nos sistemas educacionais, especialmente em contextos de desigualdade social

Guthrie et al.	2004	O sucesso escolar é alcançado quando os alunos estão envolvidos em atividades de aprendizagem significativas e têm acesso a materiais de leitura
Canário	2005	Define o sucesso escolar como a capacidade da escola se adaptar às necessidades e às expectativas dos alunos
Crosnoe & Johnson	2011	O sucesso escolar é influenciado por fatores sociais, familiares e acadêmicos, além de interações entre os estudantes e os seus pares
Justino et al.	2014	Propõem que o sucesso escolar é alcançado quando os alunos demonstram competência acadêmica e habilidades socioemocionais
Verdasca	2016	Considera que o sucesso escolar é medido pelo grau de alcance dos objetivos educativos estabelecidos, considerando tanto os resultados acadêmicos como o desenvolvimento integral do aluno
Amorim	2019	O sucesso escolar é como as estruturas escolares impactam o sucesso escolar dos alunos
Ventura	2021	Define o sucesso escolar como um conceito que incorpora o desenvolvimento integral dos alunos em várias dimensões

Fonte: Elaboração própria

Na obra de Ventura (2021), o conceito de sucesso escolar é abordado de forma a desafiar as visões tradicionais, geralmente restritas ao desempenho escolar. O autor propõe uma

abordagem mais inclusiva e equitativa, a qual não se limita às notas ou resultados em testes, mas que também considera o desenvolvimento pessoal, social e emocional dos alunos. Neste sentido, o sucesso escolar é visto como a capacitação dos estudantes para enfrentar os desafios da vida escolar e social, assegurando que todos, independentemente das suas diferenças, possam atingir o seu pleno potencial. Ao longo desta dissertação, foi considerado o conceito de sucesso escolar deste autor, ressaltando, no entanto, a dificuldade face à multiplicidade de abordagens conceptuais que existem sobre o tema.

2.2 Limitações dos indicadores tradicionais de sucesso escolar em Portugal

O presente capítulo visa explorar as limitações inerentes aos indicadores tradicionais de sucesso escolar em Portugal, reconhecendo a complexidade do sistema educativo. Ao questionar as métricas convencionais, procura-se compreender as lacunas na avaliação do desempenho escolar e identificar possíveis abordagens mais abrangentes e precisas.

Hoje em dia, existem diversas publicações sobre o (in)sucesso escolar, algumas das quais mencionadas no decorrer deste capítulo. Muitos estudos, têm adotado abordagens que, embora relevantes, podem ser complementadas com perspetivas adicionais no que diz respeito aos modelos estatísticos utilizados. Embora existam investigações de âmbito nacional que recorrem a técnicas estatísticas avançadas, é crucial incentivar estudos que explorem metodologias estatísticas variadas, de forma a introduzir novas perspetivas aprofundar a compreensão desta temática.

Em Portugal, tal como os autores referidos, muitos outros, basta procurar as dissertações e teses elaboradas sobre o tema, realizaram estudos com o intuito de apurar as possíveis causas do (in)sucesso escolar, concluindo assim que é impossível responsabilizar apenas uma entidade, visto que, a sua origem está na união de vários fatores: “condições sociais, económicas e culturais, a família, o sistema educativo e a própria escola” (Branco, 2012, citado por Martins, 2017).

Existe uma variedade de elementos responsáveis pelo (in)sucesso escolar, como refere Miranda (2010), que explica que este é o “resultado de um conjunto de fatores que atuam de modo coordenado uma vez que “nenhum deles tomado isoladamente o conseguiria provocar” (Miranda, 2010).

Partindo das obras anteriores, como as de Benavente (1990) e Roazzi e Almeida (1988), que abordaram o insucesso escolar no contexto português, torna-se evidente a necessidade de avaliar criticamente os indicadores convencionais de sucesso escolar. A visão crítica desses

estudos contribui para a compreensão de que o sucesso académico vai além de simples métricas e exige uma abordagem mais holística.

Amorim (2019) destaca fatores organizacionais como influências significativas no (in)sucesso escolar. Contudo, é crucial reconhecer que os indicadores tradicionais muitas vezes negligenciam essas dinâmicas organizacionais. A limitação em considerar as nuances da gestão escolar pode distorcer a realidade do desempenho dos alunos.

Martins (2017) abordando estratégias de prevenção e intervenção no insucesso escolar, ressalta a importância de ações preventivas. No entanto, os indicadores tradicionais podem não captar adequadamente a eficácia dessas estratégias, uma vez que se baseiam, muitas vezes, em resultados de avaliações finais medidos através, maioritariamente, de testes sem considerar o processo educacional.

A perspectiva de Mendonça (2006), focada no Arquipélago da Madeira, destaca limitações geográficas na compreensão do sucesso escolar. Os indicadores convencionais podem não ser sensíveis o suficiente para captar as disparidades regionais, negligenciando desafios específicos enfrentados por determinadas comunidades e áreas geográficas.

Entretanto, Silva et al. (2021) propõem uma abordagem qualitativa na avaliação do sucesso escolar, reconhecendo a importância de histórias individuais e contextos específicos que os indicadores tradicionais podem negligenciar.

Gomes et al. (2006) investiga como tecnologias educacionais podem fornecer informações adicionais sobre o desempenho dos alunos, expandindo as perspectivas de avaliação para além dos indicadores tradicionais. Gomes (2006) ressalta que a gravidade da situação, para além de educativa, é também social. Sabemos que a origem social se associa às taxas de insucesso e de abandono escolar, ao mesmo tempo que uma reduzida escolarização significa, também, baixa qualificação profissional, que pode comprometer o futuro dos indivíduos e das famílias (Detry & Cardoso, 1996, citado por Gomes, 2006).

A narrativa promovida pelos meios de comunicação, baseia-se no facto de que o sucesso escolar é frequentemente avaliado com base em métricas como *rankings* e desempenho em exames. No entanto, a situação torna-se realmente preocupante quando se ignora que as classificações internas dos alunos são baseadas em diversos momentos e critérios de avaliação, refletindo um processo de avaliação contínuo. É esta abordagem que as distingue das classificações externas, que visam aferir habilidades e competências que, se espera, que tenham sido ensinadas e aprendidas em certo momento da escolarização, geralmente através de avaliações pontuais e padronizadas. Posto isto, esta abordagem tem sido criticada pela sua limitação em capturar a totalidade do sucesso educativo. Autores como Afonso (2009)

argumentam que há aspetos do sucesso escolar que não são facilmente mensuráveis ou comparáveis através dessas métricas quantitativas. Além disso, pesquisadores como Almada (2015) destacam a importância das políticas sociais na promoção do sucesso educativo, enfatizando que fatores como equidade e acesso universal a qualificações são fundamentais para superar desigualdades sociais. Outros estudos, como os de Vasconcelos (2015) e Claro (2015), apontam para a influência significativa de fatores organizacionais e sociais, como cultura escolar e gestão eficaz, no sucesso dos alunos. Portanto, uma análise abrangente do sucesso escolar deve considerar uma gama mais ampla de variáveis além dos rankings e notas dos exames.

Além dos indicadores e variáveis anteriormente mencionados, geralmente utilizados na análise do sucesso escolar, é igualmente relevante discutir os modelos estatísticos que têm sido empregues na análise deste fenómeno.

De um modo geral, os modelos aplicados à análise do sucesso escolar em Portugal subdividem-se em dois tipos: estatísticos e adaptativos. No que diz respeito aos modelos estatísticos, estes baseiam-se no uso de variáveis ou indicadores para descrever a qualidade percebida ou esperada, por meio de um processamento único da informação. Já os modelos adaptativos, destacam-se por usar variáveis ou indicadores com o objetivo de identificar associações existentes, de uma forma supervisionada (com base em pressupostos ou regras) ou de uma forma não supervisionada (sem pressupostos ou regras pré-definidas).

Embora os métodos estatísticos sejam práticos, possuem algumas limitações no que diz respeito ao agrupamento das observações, pois operam somente a partir de uma perspetiva descritiva. Para compreender melhor as semelhanças das observações, é geralmente necessário analisar as variáveis a partir de uma perspetiva multiespectral, englobando múltiplas variáveis ou indicadores. Desta forma, modelos como o SOM (também conhecido como mapa de Kohonen) podem oferecer uma abordagem mais completa e vantajosa no que se refere à análise do sucesso escolar.

Ao considerar as limitações dos indicadores tradicionais de sucesso escolar, os quais se encontram inteiramente relacionados com o desempenho escolar (taxa de aprovação, retenção, abandono, conclusão, notas ou classificações), este capítulo proporciona uma visão crítica do sistema educativo em Portugal. A revisão da literatura revela a complexidade do sucesso escolar e destaca a necessidade de abordagens mais abrangentes e sensíveis aos contextos específicos. A introdução de novas perspetivas, como abordagens qualitativas e o uso de tecnologias educacionais, sugere caminhos promissores para uma análise mais precisa e abrangente do sucesso escolar no nível básico do ensino público em Portugal.

2.3 Introdução ao SOM

O avanço tecnológico e a crescente disponibilidade de dados têm proporcionado novas oportunidades para explorar e compreender padrões complexos em diversas áreas. Neste contexto, o SOM surge como uma ferramenta valiosa para análise, *clustering* e visualização de dados. Este capítulo tem como objetivo introduzir o SOM, destacando as suas características, aplicações e potencial na análise do sucesso escolar nos níveis básico do ensino público em Portugal.

Embora as redes neuronais possam ser inconstantes e aplicadas tanto em regressão como em classificação, o SOM destaca-se por ser uma técnica de aprendizagem não supervisionada, eficaz na organização e análise de dados multidimensionais.

Um neurónio apresenta capacidades como a aprendizagem, a previsão e ainda consegue memorizar. Cada input que o neurónio recebe vai, consequentemente, influenciar o *output* gerado, de acordo com o peso que cada input apresenta. Assim, as redes neuronais estão em constante aprendizagem, ao ser permitido ajustar determinado peso, com o objetivo de melhorar a performance e desempenho (Serengil & Ozpinar, 2017, citado por Narciso, 2023).

Em contrapartida, as redes neuronais podem ser inconstantes, sendo aplicadas tanto em estudos de regressão quanto em estudos de classificação. Porém, a regressão linear é um modelo utilizado em estudos de regressão, mas, diversas vezes, não lida bem com períodos temporais que não sejam lineares. De forma a combater esta situação, existem, também, alguns modelos estatísticos que podem ser aplicados em períodos temporais não lineares, como é o caso dos métodos de suavização exponencial. Ainda assim, são modelos que dependem da descoberta de possíveis tendências e do impacto sazonal nos dados a utilizar (Serengil & Ozpinar, 2017, citado por Narciso, 2023).

Para contornar as limitações dos modelos mencionadas anteriormente, surgem técnicas alternativas que permitem lidar com dados mais complexos e identificar padrões não lineares de forma mais eficaz, como é o caso do SOM.

O SOM (modelo adaptativo) proposto por Kohonen, representa uma técnica de aprendizagem não supervisionada que organiza dados multidimensionais numa estrutura bidimensional. O objetivo principal destes mapas é extrair e ilustrar as estruturas consideradas fundamentais no conjunto de dados em análise (Narciso, 2023).

Este modelo é inspirado na forma como acreditamos que o cérebro humano funciona e, na perspetiva de muitos, é um dos mais realistas ao nível do funcionamento do nosso cérebro. É considerada uma ferramenta útil para resolver problemas do mundo real e tem a capacidade de

organizar grandes conjuntos de dados através das suas semelhanças (Henriques, 2010, citado por Coelho, 2013, citado por Narciso, 2023). O autor do SOM apresenta-o como uma ferramenta analítica para a visualização e análise de grandes volumes de dados. No decurso de vários estudos, este modelo não-supervisionado, é frequentemente utilizado em várias tarefas tais como a classificação, *data mining*, *clustering*, entre outras.

O SOM é um método de rede neuronal artificial, que permite agrupar, classificar, fazer estimativas, e até mesmo prever. Este também é amplamente utilizado quando o objetivo passa por organizar aglomerados de dados (Kalteh et al., 2008, citado por Narciso, 2023). Este tipo de modelos tem como particularidade o facto de não se conhecerem nem os objetivos pretendidos nem a sua solução, visto que este modelo aprende de forma inteligente, identificando possíveis padrões, os quais são agrupados num espaço multidimensional, consoante as suas semelhanças (Narciso, 2023).

Deste modo, os neurónios que se encontrarem fisicamente mais próximos, apresentam consequentemente padrões semelhantes, aumentando a probabilidade de pertencerem ao mesmo agrupamento, pela sua similitude (Kohonen, 2001, citado por Kalteh et al., 2008, citado por Narciso, 2023).

São modelos adequados para detetar agrupamentos de padrões idênticos sem supervisão, e são muito utilizados não só como uma ferramenta analítica, mas também em diversas aplicações no mundo real. A sua forte capacidade para explorar e visualizar dados de grandes dimensões sem a necessidade de pressupostos, faz com que este modelo seja bastante atrativo para muitos investigadores (Asan & Ercan, 2012, citado por Narciso, 2023).

Negri (2013) explora os At-SOM, variantes do SOM, que apresentam características atenuantes, proporcionando uma adaptação mais flexível a diferentes conjuntos de dados. Esta introdução abordará os conceitos fundamentais para a compreensão dessas técnicas, permitindo a sua aplicação eficaz na análise do sucesso escolar.

Ultsch e Morchen (2005) oferecem *insights* sobre as diversas aplicações do SOM, destacando o seu uso em *clustering*, visualização e classificação de dados. A versatilidade dessas ferramentas é evidente, sugerindo a sua potencial aplicação na análise do sucesso escolar. A compreensão das capacidades e limitações do SOM é crucial para explorar o seu potencial de forma eficaz.

Utilizar um SOM pequeno é praticamente idêntico a uma clusterização k-Means com k igual ao número de nós no mapa. A preservação topológica da projeção do SOM é de pouca utilidade ao utilizar mapas pequenos. Fenómenos emergentes envolvem, por definição, um grande número de indivíduos, onde grande significa pelo menos alguns milhares. É por isso que

utilizamos SOMs grandes e os denominamos Mapas Auto-Organizáveis Emergentes (ESOM), para enfatizar a distinção. Foi demonstrado que utilizar ESOM é um processo significativamente diferente do uso k-Means (Ultsch e Morchen, 2005; PP. 5).

A dissertação de mestrado de Afonso (2009) ilustra também uma aplicação específica do SOM na avaliação dos parâmetros da qualidade da água. Essa abordagem prática oferece uma perspectiva concreta de como o SOM podem ser adaptados para análises específicas, indicando possíveis caminhos para a sua utilização na avaliação dos indicadores de sucesso escolar.

Considerando a literatura sobre o SOM e as suas aplicações em diferentes contextos, há uma lacuna exploratória na sua aplicação direta na área educacional, especialmente na análise do sucesso escolar. Este capítulo propõe-se a preencher essa lacuna ao introduzir o SOM como uma ferramenta promissora para a compreensão de padrões complexos nos dados educacionais.

2.4 Potencialidades do SOM na análise do sucesso escolar em Portugal

Ao abordar as potencialidades do SOM na análise do ensino em Portugal, torna-se importante explorar as capacidades inovadoras dessas ferramentas para compreender padrões educacionais complexos e proporcionar *insights* valiosos para o aprimoramento do sistema educacional.

Kohonen (2001) oferece uma visão abrangente do SOM, abordando os fundamentos teóricos e aplicações práticas. Uma leitura essencial para a compreensão aprofundada dessa técnica.

Pohjola et al (2000) explora a visualização exploratória do SOM, destacando como a representação gráfica pode enriquecer a interpretação dos resultados. Essa abordagem visual pode ser crucial na análise educacional.

A obra de Kohonen (2001) fornece *insights* sobre a aplicação do SOM na educação, destacando casos de sucesso e desafios enfrentados. Compreender as experiências anteriores é vital para adaptar essa técnica ao contexto educacional específico de Portugal.

Integrando o SOM Toolbox apresentada por Vesanto et al. (2000), este capítulo busca explorar a aplicação do SOM no contexto educacional português. A adaptação desta ferramenta para considerar características específicas, como o currículo português e as políticas educacionais, é crucial.

A visualização do SOM, enfatiza a importância dessa abordagem na identificação de padrões. Ao aplicar essa técnica na análise do ensino em Portugal, a visualização dinâmica pode oferecer uma compreensão mais rica dos dados educacionais.

Apesar de não ser diretamente relacionado à educação, Franco et al. (2002) aborda a motivação dos trabalhadores do setor público, oferecendo uma perspectiva sobre como fatores motivacionais podem ser incorporados na análise de indicadores de sucesso escolar.

A obra de Franco et al. (2002) contribui para a discussão ao considerar fatores motivacionais na análise de indicadores. Ao aplicar essa abordagem à educação, o SOM pode ser empregue para explorar a relação entre motivação, desempenho escolar e outros fatores relevantes.

Oferecendo estratégias de ensino baseadas em pesquisas e proporcionando ideias para o desenvolvimento de estratégias educacionais personalizadas, a integração destas estratégias com o SOM pode potencializar a eficácia das intervenções educacionais.

Além das referências fornecidas, a obra de Marzano et al. (2001) apresenta uma visão sobre estratégias de ensino com base em pesquisas, sugerindo possíveis perspectivas futuras na análise do ensino em Portugal. A compreensão dessas estratégias é crucial para superar desafios e garantir a eficácia das intervenções baseadas no SOM.

Importa reter que o SOM tem o potencial de contribuir significativamente para a análise do sucesso escolar em Portugal. Essas ferramentas oferecem uma abordagem inovadora para compreender padrões educacionais complexos e identificar fatores que influenciam o desempenho dos alunos. Em primeiro lugar, o SOM pode ajudar a identificar padrões de desempenho escolar em diferentes regiões do país, i.e. por NUTS III², permitindo uma compreensão mais detalhada das disparidades educacionais. Além disso, esses mapas podem explorar uma variedade de fatores que têm impacto no sucesso dos alunos, como recursos escolares, características socioeconómicas dos alunos e políticas educacionais locais. A capacidade de visualizar dados de forma dinâmica e gráfica oferecida pelo SOM facilita a interpretação dos resultados e permite identificar tendências temporais no desempenho escolar. Com base nessas análises, é possível desenvolver intervenções educativas personalizadas e direcionadas para promover o sucesso dos alunos, adaptando estratégias de ensino e implementando programas de apoio académico conforme necessário. Em suma, o SOM representa uma ferramenta poderosa para compreender e melhorar o sucesso escolar em Portugal, fornecendo dados para os formuladores de políticas educacionais, pesquisadores e profissionais da educação.

Este capítulo proporcionou uma exploração detalhada das potencialidades do SOM na análise do ensino em Portugal, incorporando novas referências para ampliar a compreensão. A

² As NUTS III dividem-se em: Norte, Centro, Lisboa e Vale do Tejo, Alentejo e Algarve.

próxima fase da pesquisa será dedicada à aplicação prática dessas ferramentas inovadoras, considerando as nuances do sistema educacional português com o objetivo de contribuir com possíveis melhorias significativas.

Capítulo 3. Metodologia de Investigação

Neste capítulo é abordada a metodologia adotada ao longo desta dissertação. Assim, são destacados os dados utilizados na análise do tema questão, assim como o modelo e a linguagem empregados na realização dos objetivos definidos.

3.1 Metodologia adotada

Existem diversas técnicas utilizadas para exploração de dados e desenvolvimento de modelos. As três técnicas mais abrangentes e amplamente utilizadas na análise exploratória de dados são CRISP-DM, SEMMA e KDD.

O SEMMA (*Sample, Explore, Modify, Model e Asses*), desenvolvido pelo SAS Institute, é uma metodologia cíclica, aplicada para executar tarefas de *data mining*, desde a sua extração até à avaliação dos modelos construídos. Esta técnica é muito idêntica ao CRISP-DM, porém não se foca na compreensão do negócio e nas suas particularidades.

A metodologia KDD (*Knowledge Discovery from Databases*), também conhecida por descoberta de conhecimento através de bases de dados, surge na década de 1990 por Gregory Piatetsky-Shapiro e Usama Fayyad. Esta técnica permite identificar padrões ao longo dos dados, criando, assim, conhecimento a partir dos mesmos.

O CRISP-DM (*Cross Industry Standard Process for Data Mining*) é uma das técnicas mais utilizadas em *data mining*. Muitos projetos dentro da área de aplicações *model-driven* são frequentemente realizados de maneira exploratória e não estruturada, e para dar suporte ao desenvolvimento deste tipo de aplicações, foram criados modelos e estruturas de processo de desenvolvimento, como o CRISP-DM (Grimheden & Järlesäter, 2022, citado por Narciso, 2023).

Silva (2021), argumenta que a metodologia CRISP-DM oferece várias vantagens, como a capacidade de se adaptar a qualquer tipo de projeto, uma vez que não depende de uma única ferramenta para ser implementada, e a sua natureza cíclica, que permite avançar ou retroceder no processo a qualquer momento.

O planeamento das organizações é facilitado pelas diversas metodologias que estão, normalmente, concebidas para ajudar no desenvolvimento de processos, de cariz analítico. Algumas dessas metodologias são estruturadas para trabalhar com determinadas ferramentas de *software* específicas, no entanto, até ao momento, a metodologia *open-standard* mais popular é o CRISP-DM (Shearer, 2000, citado por Quinn, 2020, citado por Narciso, 2023).

De modo a sustentar esta dissertação, e sendo a análise do sucesso escolar uma temática em *Business Analytics*, a metodologia adotada no decorrer deste estudo é a metodologia CRISP-DM.

3.2 Metodologia CRISP-DM

O subcapítulo que se segue aborda as diferentes etapas desta metodologia cíclica. Posto isto, serão abordadas as suas seis etapas, conforme ilustrado na Figura 1.

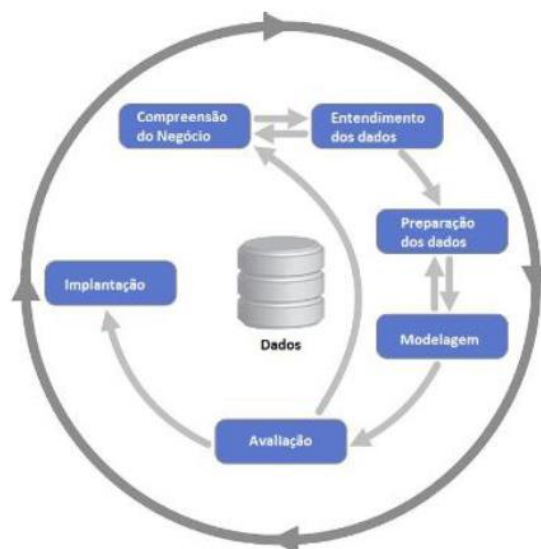


Figura 1- Diagrama do modelo CRISP-DM (Ramos, et al., 2020)

3.2.1 Compreensão do negócio

A compreensão do negócio é o ponto de partida desta metodologia. Esta primeira fase consiste no conhecimento do negócio e na compreensão dos objetivos do projeto e é a fase mais importante deste modelo (Shearer, 2000, citado por Ramos et al., 2020, citado por Narciso, 2023). Segundo Grimheden e Järlesäter (2022), esta fase permite compreender quais são os requisitos do modelo de dados numa perspetiva de negócio.

No caso da presente dissertação, a compreensão do negócio foi explorada através da compreensão do conceito de sucesso escolar, especialmente da sua evolução ao longo dos anos, e da identificação das variáveis relevantes, tendo em conta as dimensões essenciais à análise do sucesso escolar.

Da abordagem destas temáticas no decorrer dos capítulos 1 e 2, resultaram como principais ideias para a compreensão do negócio que:

- O sucesso escolar é o resultado de diversos fatores que, quando combinados, compõem uma experiência educacional mais alinhada com a realidade;
- Existem dados disponíveis sobre três dimensões específicas: socioeconómicas, organizacionais, e de sucesso escolar – e, como a literatura aponta para a sua importância, optou-se por focar nestas;
- Associadas a estas três dimensões existem 33 variáveis de base inerentes à análise do sucesso escolar, que se apresentam detalhadas no subcapítulo seguinte;
- A análise do sucesso escolar é fundamental para identificar áreas de melhoria, orientar políticas educativas eficazes e promover o desenvolvimento integral dos alunos.

O acompanhamento e o apoio aos alunos são amplamente reconhecidos como fatores essenciais por diversos agentes. A excelência no ensino é frequentemente considerada uma vantagem competitiva das escolas, contribuindo de forma decisiva para o sucesso escolar no ensino básico em Portugal.

3.2.2 Compreensão dos dados

A compreensão dos dados é a segunda etapa desta metodologia. Segundo Quinn (2020), esta fase tem início com a recolha e descrição inicial dos dados. Além disso, inclui a exploração dos dados, com o objetivo de compreender as suas inter-relações, avaliar a sua qualidade e, por fim, identificar possíveis problemas associados.

A dissertação realizada teve por base dados recolhidos da Direção Geral de Estatísticas da Educação e Ciência (DGEEC), que se encontravam desagregados por escola. A recolha de dados foi orientada por indicadores e variáveis que, em publicações anteriores, se revelaram cruciais para a análise do sucesso escolar no ensino básico público português. Estes dados incluem informações relacionadas com fatores socioeconómicos, organizacionais e outros que afetam o sucesso escolar. A recolha de dados foi realizada no dia 9 de setembro no safe center da DGEEC.

Para a sua elaboração, e após várias pesquisas sobre o tema, tem-se uma grelha de avaliação (2024), presente na Tabela 2 e mais detalhada no Anexo A, que contempla as diversas dimensões e variáveis identificadas ao longo dos capítulos anteriores.

Esta grelha foi essencial para garantir a consistência na interpretação dos dados e assegurar que cada variável fosse adequadamente enquadrada no contexto da análise do sucesso escolar.

Neste sentido, as dimensões escolhidas para o efeito foram três:

- Socioeconómicas: compreende as condições financeiras, educativas e profissionais das famílias dos alunos, assim como a sua diversidade cultural;
- Organizacionais: diz respeito à composição e estabilidade do corpo docente, bem como à distribuição e localização das escolas;
- Sucesso escolar: comporta o desempenho escolar e a progressão dos alunos, refletindo as taxas de retenção, desistência e a continuidade no percurso escolar.

Avaliar o contexto socioeconómico é fundamental numa dissertação que incide sobre o sucesso escolar, pois este apresenta uma influência direta no desempenho escolar dos alunos, como evidenciado pelos estudos citados nos capítulos de enquadramento. Aspetos como o acesso a apoios sociais, as necessidades específicas de educação, a origem, a escolaridade, o grupo etário e a profissão dos pais e encarregados de educação afetam o ambiente familiar, as oportunidades de aprendizagem e o suporte que os alunos recebem.

Na análise do sucesso escolar, é importante avaliar o contexto organizacional, na medida em que a estrutura e os recursos humanos de uma escola influenciam diretamente a qualidade do ensino e a aprendizagem dos alunos, como evidenciado pelos estudos citados nos capítulos de enquadramento. Fatores como a proporção de docentes efetivos e contratados, o seu grupo etário predominante e as habilitações académicas refletem a estabilidade e continuidade do corpo docente, aspetos que impactam o ambiente de ensino. A distribuição etária dos professores pode influenciar a experiência e capacidade de adaptação pedagógica.

Por fim, a última dimensão eleita refere-se ao sucesso escolar, a qual mede diretamente o desempenho escolar e a progressão dos alunos ao longo do ensino. A média das provas nacionais reflete o nível de aprendizagem e competência dos estudantes, enquanto as taxas de retenção e desistência indicam os desafios no percurso escolar, como dificuldades de aprendizagem ou contextos adversos que levam ao insucesso. Os percursos diretos de sucesso mostram a capacidade dos alunos de progredirem sem interrupções, e a média de retenção por aluno permite avaliar a repetência acumulada.

A Tabela 2 ilustra a grelha de avaliação resumida, que contem as diversas variáveis utilizadas para analisar o sucesso escolar dos alunos no ensino básico do sistema público em Portugal.

No Anexo A, é possível encontrar uma versão mais detalhada dessa tabela, onde estão discriminadas todas as variáveis mencionadas, juntamente com as suas definições. Essa informação adicional permite uma visão mais abrangente das dimensões utilizadas, facilitando

a interpretação dos resultados e a compreensão do impacto de cada variável na análise do sucesso escolar.

Tabela 2- Grelha de avaliação resumida

Designação da escola:		Ano letivo:
Região:		Distrito:
Dimensões	Sub-dimensões	Variáveis
Socioeconómicas	Composição dos alunos	Percentagem de alunos com ASE
		Percentagem de alunos com PEI
		Percentagem de alunos com nacionalidade estrangeira
	Contexto social	Percentagem de alunos com pais e EE com nacionalidade estrangeira
		Grupo etário predominante dos pais e EE
		Percentagem de pais e EE com ensino secundário ou superior completo
		Percentagem dos tipos de situação profissional dos pais ou EE
		Tipos de profissões dos pais ou EE
Organizacionais	Estrutura da escola e suporte	Acesso à internet
		Acesso a apoio psicológico
		Número de turmas por ciclo
	Distribuição dos alunos	Género predominante dos alunos
		Percentagem de alunos no 1º ciclo
		Percentagem de alunos no 2º ciclo
		Percentagem de alunos no 3º ciclo
	Corpo docente	Percentagem de docentes do quadro
		Percentagem de docentes contratados
		Número de docentes no ensino básico
		Grupos etários predominante dos docentes
		Habilitação predominante dos docentes
		Média de tempo de serviço dos docentes
Sucesso escolar	Desempenho escolar	Média das provas nacionais e exames
		Taxa de retenção
		Taxa de desistência

		Percursos Diretos de Sucesso
--	--	------------------------------

Fonte: Elaboração própria (com base em Dawes, 2000; Guerreiro, 2016; Naik e Gantasala, 2010; Seenivasan e Anbazhag, 2014)

Após a elaboração da grelha acima representada, foi possível dar início à recolha dos dados, de forma a obter a informação necessária para as etapas subsequentes do CRISP-DM. A grelha de avaliação foi aplicada a 3 914 escolas que ministram o ensino básico do sistema público português, nos anos letivos entre 2019/2020 e 2021/2022. Neste sentido a amostra recolhida totaliza 11 742 registos³. A Tabela 3 apresenta as diferentes variáveis incluídas na grelha de avaliação e especifica o tipo de variável em causa.

Tabela 3- Tipos de variáveis presentes na grelha de avaliação

Variável	Designação	Tipo de variável
ID_escola	ID da escola	Numérica
Desig_escola	Designação da escola	Texto
Regiao	Região	Texto
Distrito	Distrito	Texto
Acesso_internet	Acesso à internet	Texto
Acesso_psicologo	Acesso a apoio psicológico	Texto
N_turma_ciclo	Número de turmas por ciclo	Numérica
Perc_alunos_ASE	Percentagem de alunos com ASE	Numérica
Perc_alunos_PEI	Percentagem de alunos com PEI	Numérica
Perc_alunos_estrang	Percentagem de alunos com nacionalidade estrangeira	Numérica
Genero_pred_alunos	Género predominante dos alunos	Texto
Perc_alunos_1ciclo	Percentagem de alunos no 1º ciclo	Numérica
Perc_alunos_2ciclo	Percentagem de alunos no 2º ciclo	Numérica
Perc_alunos_3ciclo	Percentagem de alunos no 3º ciclo	Numérica
Perc_alunos_pais_estrang	Percentagem de alunos com pais e EE com nacionalidade estrangeira	Numérica
Grupo_etario_pred_pais	Grupo etário predominante dos pais e EE	Texto
Perc_alunos_pais_ensino	Percentagem de pais ou EE com ensino secundário ou superior completo	Numérica

³ Corresponde a 3 anos letivos (3*3 914= 11 742)

Perc_sit_prof_trab_outrem	Percentagem de pais ou EE que trabalham por conta de outrem	Numérica
Perc_sit_prof_trab_prop	Percentagem de pais ou EE que trabalham por conta própria	Numérica
Perc_sit_prof_trab_desemp	Percentagem de pais ou EE que estão desempregados	Numérica
Perc_sit_prof_trab_reform	Percentagem de pais ou EE que estão reformados	Numérica
Tipo_prof_pais	Tipos de profissões dos pais ou EE	Texto
Perc_doc_quadro	Percentagem de docentes do quadro	Numérica
Perc_doc_contratado	Percentagem de docentes contratados	Numérica
N_doc_básico	Número de docentes no ensino básico	Numérica
Grupo_etario_pred_doc	Grupos etários predominante dos docentes	Texto
Hab_pred_doc	Habilitação predominante dos docentes	Texto
Med_tempo_servico_doc	Média de tempo de serviço dos docentes	Numérica
Med_provas_exames	Média das provas nacionais e exames	Numérica
Taxa_retencao	Taxa de retenção	Numérica
Taxa_desistencia	Taxa de desistência	Numérica
Perc_diretos	Percursos Diretos de Sucesso	Numérica
Ano_letivo	Ano letivo	Numérica

Fonte: *Elaboração própria*

3.2.3 Preparação dos dados

A preparação dos dados é a terceira etapa da metodologia CRISP-DM. Esta fase abrange as várias ações necessárias para construir o reporte de dados final. Esses dados serão trabalhados e preparados cuidadosamente, a fim de se obter uma base de dados consolidada, fazendo parte dela a limpeza, construção de dados e dos seus atributos, e ainda a sua formatação (Shearer, 2000, citado por Ramos et al., 2020; Quinn, 2020, citado por Narciso, 2023).

Para garantir melhores *outputs* e uma análise mais eficaz, as variáveis presentes na Tabela 3 foram analisadas, com o objetivo de avaliar a necessidade de serem alteradas, adaptadas, eliminadas e/ou transformadas.

Inicialmente foi definido que as escolas correspondem à variável “ID_escola” através do recurso à ferramenta Python. De seguida, as variáveis “Desig_escola”, “Distrito” e “Ano_letivo” foram excluídas, não tendo sido consideradas para o modelo (*vide script* Python na Figura 2). A variável *Desig_escola* foi eliminada devido à redundância, uma vez que já existe

uma outra variável (ID_escola) que fornece a mesma informação. Neste seguimento, foi também eliminada a variável *Distrito*, pois evita redundância, simplifica a análise, visto que a localização por Região torna-se menos fina, e ajuda a evidenciar padrões mais amplos de sucesso escolar e a variável *Ano_letivo*, visto que pode introdução ruído e complexidade desnecessária.

```
df.drop(columns=['Desig_escola', 'Distrito', 'Ano_letivo'], inplace = True)
df.head()
```

Figura 2- Código- Eliminação das variáveis em Python

Dado que estamos na presença de linguagem de programação, é comum que este tipo de modelos não utilizem dados categóricos em formato de texto. Por esse motivo, e considerando que as variáveis categóricas analisadas não apresentavam qualquer relação ordinal entre elas, foi aplicada a função *get_dummies* para converter as variáveis categóricas, presentes na grelha de avaliação, em variáveis *dummies*, conforme demonstra a Figura 3. Esta transformação garante que estas variáveis possam ser utilizadas na modelação, fase seguinte do CRISP-DM.

```
df = pd.get_dummies(df, columns=['Região'], prefix=['Região_'])
df = pd.get_dummies(df, columns=['Acesso_internet'], prefix=['Acesso_internet_'])
df = pd.get_dummies(df, columns=['Acesso_psicólogo'], prefix=['Acesso_psicólogo_'])
df = pd.get_dummies(df, columns=['Gênero_pred_alunos'], prefix=['Gênero_pred_alunos_'])
df = pd.get_dummies(df, columns=['Grupo_etário_pred_pais'], prefix=['Grupo_etário_pred_pais_'])
df = pd.get_dummies(df, columns=['Tipo_prof_pais'], prefix=['Tipo_prof_pais_'])
df = pd.get_dummies(df, columns=['Grupo_etário_pred_doc'], prefix=['Grupo_etário_pred_doc_'])
df = pd.get_dummies(df, columns=['Hab_pred_doc'], prefix=['Hab_pred_doc_'])
df.head()
```

Figura 3- Código- Transformação de variáveis categóricas em dummies em Python

Após a exclusão e formatação das variáveis acima mencionadas, foi aplicado o comando *StandardScaler* que permite a estandardização das variáveis, removendo a sua média e a variância unitária. Em seguida, todas as variáveis foram normalizadas, de modo a garantir que todas beneficiassem do mesmo peso, evitando que alguma apresentasse um peso maior que outra no SOM. Estas operações estão representadas na Figura 4.

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df_normalized = scaler.fit_transform(df)
df = pd.DataFrame(df_normalized, columns=df.columns)
df.head()
```

Figura 4- Código- Estandarização e normalização das variáveis selecionadas em Python

A grelha de avaliação continha *missing values* (na) em algumas variáveis, conhecidos como dados omissos. Para combater esta lacuna, foi utilizada a função *fillna*, que preenche os campos vazios. Neste caso específico, os *missing values* foram substituídos pelo valor zero, como ilustra a Figura 5.

```
df.fillna(0, inplace=True)
```

Figura 5- Código- Preenchimento de missing values em Python

A Tabela 4 apresenta um resumo das variáveis que sustentaram algum tipo de modificação ou formatação durante as etapas de compreensão e preparação dos dados recolhidos.

Tabela 4- Síntese das variáveis transformadas

Variável	Tipo de variável	Tipo de transformação
Desig_escola	Texto	Eliminada
Distrito	Texto	Eliminada
Ano_letivo	Numérica	Eliminada
Região	Texto	Numérica
Acesso_internet	Texto	Numérica
Acesso_psicologo	Texto	Numérica
Genero_pred_alunos	Texto	Numérica
Grupo_etario_pred_pais	Texto	Numérica
Tipo_prof_pais	Texto	Numérica
Grupo_etario_pred_doc	Texto	Numérica
Hab_pred_doc	Texto	Numérica

Fonte: Elaboração própria

Com base nas etapas descritas anteriormente, e considerando a transformação e eliminação de variáveis não essenciais ao modelo selecionado, a Tabela 5 exibe o processo de triagem aplicado à amostra inicial de dados. Essa amostra inicial inclui as variáveis relacionadas à avaliação das diversas escolas. Após este processo, o número de avaliações por observação é incluído na etapa seguinte – a modelação.

Tabela 5- Processo de seleção da amostra

Descrição	Nº de observações
Dados recolhidos através da grelha de avaliação	387.486 ⁴
Após a eliminação das variáveis desnecessárias ao modelo	340.547

Fonte: Elaboração própria

3.2.4 Modelação

A etapa da modelação é a fase que envolve atividades como a seleção de técnicas de modelação e calibração dos seus parâmetros. Usualmente, existem várias técnicas diferentes que podem ser usadas para um único problema de prospeção de dados. Importa, também, considerar o caso de uso específico e escolher o algoritmo que será o mais adequado para o efeito (Grimheden & Järlesäter, 2022, citado por Narciso, 2023). Ramos et al. (2020), reconhecem que na modelação de dados são escolhidas as técnicas de modelação adequadas, em conformidade com algoritmos de *data mining* e posterior realização de testes para ajustar os parâmetros desses mesmos algoritmos. Aqui, é comum regressar à etapa anterior, pois, por vezes, é necessário ajustar e manipular os dados (Narciso, 2023).

Conforme discutido no Capítulo 2- Revisão da Literatura, e tendo em atenção a natureza de dados que seriam recolhidos e, posteriormente, processados, considera-se que o modelo escolhido para a modelação dos dados foi o modelo não supervisionado SOM. Este método, como o próprio nome sugere, é auto-organizado e amplamente utilizado em situações em que se desconhece o problema e, à priori, o número de agrupamentos. Este estudo enquadra-se nesse contexto, focando na identificação de grupos e padrões com proximidade e similitude, com a finalidade de analisar o sucesso escolar no ensino básico público e orientar as decisões de intervenção para melhorar a performance das escolas em análise.

O SOM tem a particularidade de agrupar observações de acordo com as suas semelhanças, e por esse motivo, no mapa topológico pode surgir a mesma escola em grupos separados, uma vez que foram realizadas 3 observações por escola, i.e. entre os anos letivos 2019/2020 e 2021/2022. A maior parte da preparação de dados, assim como esta etapa, foi realizada utilizando a linguagem de programação Python. O código aplicado para a modelação é *open source*, e está disponível através da hiperligação <https://github.com/JustGlowing/minisom>. O código utilizado integra o MiniSom, uma versão mais simples do SOM, concebida

⁴ O número de observações é calculado a partir da multiplicação do número de variáveis (33) pelo número de escolas no total dos três anos letivos (11 742)

especialmente para estudantes como principais utilizadores, garantindo maior autonomia e compreensão sobre o funcionamento do modelo.

Na versão do MiniSom acessível através do *link* citado, foi incorporada a funcionalidade do mapa de cores, também designado *colormap*, com o intuito de substituir o mapa inicial por um mapa de cores padrão.

Além disto, este gráfico possibilita, ainda, a obtenção das distâncias médias entre os neurónios para todos os agrupamentos vencedores. A Figura 6 exibe o modelo implementado, em linguagem de programação, e nela estão representados os vários parâmetros suscetíveis de alteração, de modo a alcançar um dos objetivos principais – a identificação de agrupamentos e distâncias entre eles.

```
from minisom import MiniSom
import numpy as np
import matplotlib.pyplot as plt
from pylab import bone, pcolor, colorbar, plot, show

data=df.values
n=?
som = MiniSom(x=n, y=n, input_len=len(data[0]), sigma=?,
learning_rate=?, neighborhood_function='gaussian', random_seed=0)

som.random_weights_init(data)
som.train_random(data = data, num_iteration = ?)

bone()
pcolor(som.distance_map().T)
colorbar()
for i, x in enumerate(data):
    w = som.winner(x)
    plot(w[0] + 0.5,
         w[1] + 0.5,
         'o',
         markeredgecolor = 'r',
         markerfacecolor = 'r',
         markersize = 4,
         markeredgewidth = 2)
show()
```

Figura 6- Código- Aplicação do MiniSom em Python

As etapas de modelação associadas ao MiniSom, seguem a lógica de treinamento iterativa do SOM e consistem nas seguintes fases:

- Primeira etapa – cálculo aleatório dos pesos das observações de cada neurónio;
- Segunda etapa - determinação do número de iterações e cálculo gradual da distância entre cada observação e o vetor dos pesos, tendo em conta a distância euclidiana;

- Terceira etapa - comparação da distância euclidiana com o peso de cada observação para identificar o neurónio com a menor distância, denominado BMU (*Best Matching Unit*), que será o neurónio a receber um treinamento mais intenso;
- Quarta etapa - realização do treino de todos os neurónios a partir do BMU, com atualização dos pesos de cada neurónio da rede tendo em conta a função de treinamento (função de vizinhança).

O parâmetro “n” representa a dimensão da matriz gerada pelo SOM. O número de nós que serão delimitados será o número máximo de *clusters* que o SOM pode formar (Ali & Amin, 2019, citado por Narciso, 2023). O sigma é outro dos parâmetros que deve ser preenchido. Este parâmetro refere-se ao raio e agrupa os neurónios mais próximos do neurónio onde a função de ativação está localizada. Por padrão, varia entre 5% a 15% do parâmetro do sigma. Neste código específico, a função de ativação pré-definida é a gaussiana, comumente referida como *mexican hat*. Essa função transmite informação resultante da combinação linear das entradas e dos respetivos pesos, ou seja, transmite a informação por meio dos vetores de saída. Outro parâmetro a ser considerado é o *learning rate*, que permite regular o tamanho dos vetores de peso. O seu valor varia entre 0 e 1, e determina a magnitude com que os pesos dos neurónios são alterados em cada iteração, à medida que estes se aproximam do mínimo de uma função. Para este modelo foi definido um *learning rate* de 0,07, pois quanto mais baixo este parâmetro for, mais facilmente se alcança um ajuste adequado dos vários agrupamentos. Um *learning rate* elevado vai fazer com que as escolas se agrupem ao *cluster* mais próximo, em vez de ser ao mais adequado. As iterações são outro fator importante, pois indicam o número máximo de iterações por amostra. A função usada no modelo permitiu treinar o SOM de forma sequencial, passando por todos os vetores de dados. No que diz respeito à escolha do número de agrupamentos, a análise inicial foi realizada a partir da matriz 5x5, pois foi a partir daqui que se observou uma desconexão em relação à sua semelhança com a matriz. A partir deste ponto explorou-se os possíveis cenários de agrupamento, tendo-se estabelecido como limite de visualização a matriz 15x15, com base na avaliação dos parâmetros de qualidade obtidos (*vide* subcapítulo 3.2.5). A Tabela 6 apresenta os diferentes cenários de agrupamento para cada parâmetro do modelo, com o objetivo de selecionar o cenário ideal.

Tabela 6- Conjugação dos parâmetros do SOM

Nº de <i>clusters</i> (n)	Sigma	<i>Learning rate</i>	Nº de iterações
5	0,2	0,07	2000

6	0,3	0,07	2000
7	0,3	0,07	2000
8	0,4	0,07	2500
9	0,4	0,07	2500
10	0,5	0,07	2500
11	0,5	0,07	2500
12	0,6	0,07	3000
13	0,6	0,07	3000
14	0,7	0,07	3000
15	0,7	0,07	3000

Fonte: Elaboração própria

Neste tipo de análise torna-se importante verificar a evolução dos parâmetros do SOM em função do aumento do número de *clusters*, permitindo ajustar a desagregação dos agrupamentos. Esta variação de configurações é essencial para garantir que os resultados sejam otimizados, oscilando precisão e complexidade. Os resultados destes agrupamentos serão explorados no capítulo seguinte (Capítulo 4- Resultados e Discussão) para avaliar o cenário ideal, tal como foi mencionado anteriormente.

3.2.5 Avaliação dos resultados

Para Quinn (2020), a principal tarefa associada à fase de avaliação do modelo é avaliar os resultados da fase da modelação, isto é, avaliar o modelo construído de acordo com o problema formulado inicialmente na primeira fase da metodologia, a compreensão do negócio. Importa, assim, garantir a qualidade do modelo, tanto do ponto de vista da análise dos dados, como também do ponto de vista do negócio (Grimheden & Järlesäter, 2022, citado por Narciso, 2023).

Segundo Forest et al. (2020), a performance do SOM pode ser medida e avaliada, sendo que essa avaliação pode variar conforme o objetivo desejado. As métricas mais utilizadas estão agrupadas em duas principais categorias: as métricas de agrupamento (*clustering metrics*) e as métricas topográficas (*topographic metrics*).

As métricas de agrupamento são baseadas unicamente nos vetores de padrão, sem considerar a organização topológica. Por outro lado, as métricas topográficas têm em conta a topologia do mapa, especificamente do SOM, e apresentam a capacidade de identificar anomalias na vizinhança.

Para além destas duas famílias, cada uma delas divide as suas métricas em conformidade com o cálculo dos índices, i.e, se são índices internos ou externos (Forest, 2020, citado por Narciso, 2023). Considerando as diversas métricas de avaliação do desempenho do modelo, as selecionadas para avaliar a performance do SOM foram:

- Erro de quantização (*Quantization error*) – calcula o erro médio ao gerar o *output* do SOM, utilizando a distância euclidiana. Por outras palavras, mede a distância média euclidiana entre uma amostra de dados e o seu neurónio mais próximo. Quanto menor o erro de quantização, melhor será a qualidade do modelo;
- Erro topográfico (*Topographic error*) – calcula a fração de amostras para as quais os dois melhores neurónios não são vizinhos no mapa. Esse erro quantifica a suavidade do delineamento do SOM. Quanto menor for o erro topológico, maior é a preservação da topologia do mapa.

Com o intuito de ilustrar as métricas selecionadas, as quais medem a performance de cada mapa, a Figura 7 apresenta o *script* relativo ao cálculo das duas métricas mencionadas, também com recurso a linguagem de programação.

```
print("\n\nquantization error:",som.quantization_error(data),"\n")
print("topographic error:",som.topographic_error(data),"\n\n\n")
```

Figura 7- Código- Aplicação das métricas de avaliação da performance em Python

3.2.6 Implementação

A implementação do modelo, exhibe-se como a última etapa desta metodologia. Esta é a fase onde se decide a utilização ou não do modelo, e se disponibilizam os resultados ao utilizador final (Silva, 2021, citado por Narciso, 2023). Dependendo do projeto em questão, esta etapa pode não existir, pois pode apenas exigir a criação do modelo e não a sua implementação (Ramos et al., 2020, citado por Narciso, 2023). Posto isto, esta etapa permite abordar as questões de investigação apresentadas, proporcionando o desenvolvimento e a exploração aprofundada nos capítulos seguintes. No contexto desta dissertação, o modelo foi utilizado para sistematizar os agrupamentos das observações e avaliar o nível de desempenho do sucesso escolar no ensino básico, com o objetivo de definir políticas educativas de intervenção eficazes para promover o desenvolvimento dos alunos nas escolas em análise. A Figura 8 apresenta o processo de análise realizado a partir dos resultados obtidos pelo SOM.

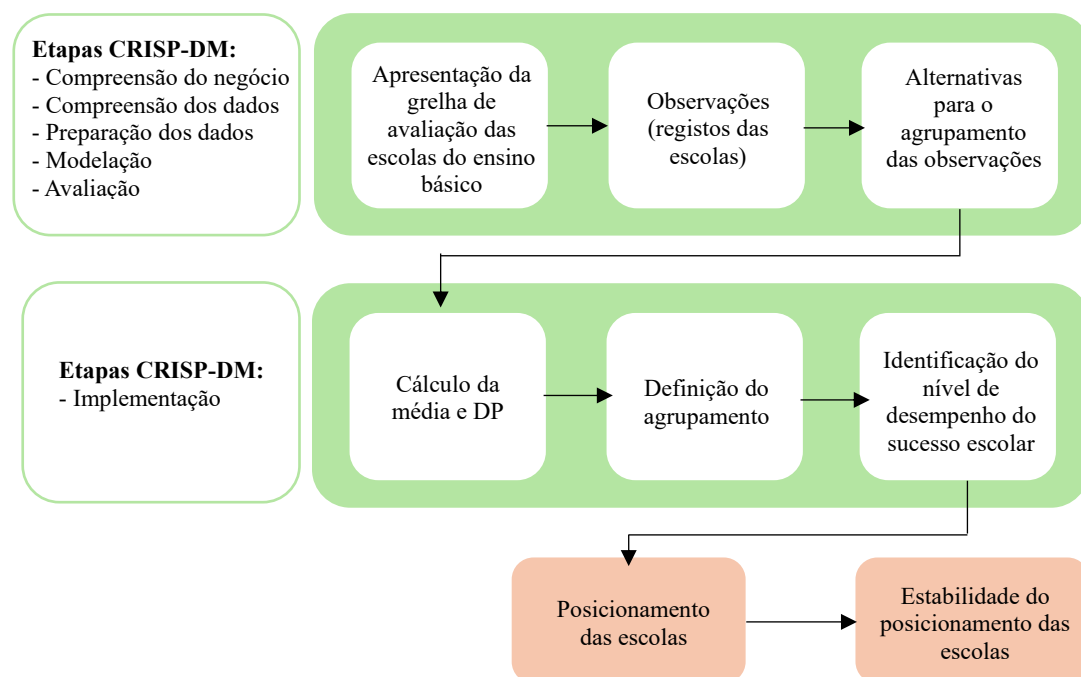


Figura 8- Processo de implementação dos resultados obtidos no SOM (Adaptado de Narciso, 2023)

As cinco primeiras etapas do CRISP-DM estão representadas na primeira parte da figura. A partir dos *outputs* gerados pelo modelo, procede-se, em seguida, à seleção do agrupamento e à definição do nível de desempenho, com base no qual foi analisado o posicionamento das escolas e a sua estabilidade ao longo dos anos letivos. O nível de desempenho da qualidade foi calculado para os agrupamentos formados, utilizando a média das 30 variáveis.

Capítulo 4. Resultados e Discussão

O Capítulo 4 tem como objetivo dar continuidade ao capítulo anterior – Metodologia. Após ter sido apresentada a metodologia, as etapas da mesma, e ainda o modelo utilizado para a concretização dos objetivos desta dissertação, este capítulo permite aplicar os resultados do SOM e analisar as questões de investigação inicialmente propostas.

4.1 Compreensão da complexidade do problema

O SOM é um modelo que apresenta grandes vantagens no agrupamento e visualização de observações com comportamentos semelhantes.

Não existindo conhecimento, *a priori*, da complexidade do problema, i.e., dos resultados da análise do sucesso escolar nas escolas básicas do ensino público em Portugal, foi necessário

inicialmente compreender a realidade existente. Ou seja, se nos deparamos com um painel de observações altamente diversificado, com comportamentos muito distintos, ou se, pelo contrário, existe uma maior proximidade no comportamento das observações, o que torna a situação mais desafiante.

Esta análise foi realizada com base em medidas descritivas, nomeadamente a média e o desvio-padrão. A utilização destas medidas é particularmente útil neste caso, visto que ambas oferecem uma visão abrangente do comportamento dos dados. Outras medidas, como a mediana ou a amplitude interquartil, embora pertinentes em casos de distribuições assimétricas ou com *outliers*, não captam de forma tão precisa a dispersão geral dos dados em relação ao ponto central. Assim, a combinação da média e do desvio-padrão revelam-se mais adequados para compreender a diversidade ou a semelhança dos comportamentos observados.

A média diz respeito à concentração dos dados, ou seja, permite verificar onde se encontra uma maior concentração de um valor específico. Trata-se de uma medida central que facilita a interpretação dos dados de uma amostra e é uma das ferramentas estatística mais utilizadas.

Relativamente ao desvio-padrão, por sua vez, é uma medida de dispersão que indica o quão dispersos os dados estão em relação à média amostral. Quando o desvio-padrão é baixo, significa que os dados da amostra estão, vulgarmente, próximos do valor médio. Em contrapartida, um desvio-padrão elevado indica que os dados da amostra estão mais dispersos, variando significativamente em torno da média.

A Tabela 7 apresenta os valores destas duas medidas tendo por base o total das observações obtidas. Além disto é ainda possível analisar na tabela as observações repartidas pelos três anos letivos analisados. Os dados em análise totalizam 11 742 registos realizados às 3 914 escolas, com base numa grelha de avaliação composta por 30 variáveis.

Tabela 7- Média e desvio-padrão das observações obtidas

Amostra	Média	Desvio-padrão	Total de observações
2019/2020	0,450	0,000	3 914
2020/2021	1,000	0,014	3 914
2021/2022	0,000	1,000	3 914
Global (3 anos letivos)	0,483	0,804	11 742

Fonte: Elaboração própria

De um modo geral, a média das avaliações do sucesso escolar é uma média baixa, numa escala de 0 a 1 (0,483). Face aos três anos letivos, existem valores dispersos, com o valor mais

alto a ser 1,0 (2020/2021) e o mais baixo a ser 0,0 (2021/2022). No entanto, constata-se que o ano letivo 2019/2020 apresenta uma média intermediária (0,450).

Tendo em conta os valores representados na tabela anteriormente mencionada, a qualidade obtida através da experiência (percebida) é intermédia.

O desvio-padrão da amostra total, também calculado, é considerado elevado. Isto indica que as diversas observações apresentam uma tendência dispersa em relação ao valor médio calculado. O ano letivo 2021/2022 foi o ano com um desvio-padrão mais elevado (1,000), no entanto, os restantes anos apresentam valores baixos. Estes valores de desvio-padrão permitem concluir que as várias observações apresentam valores um pouco dispersos em relação à média amostral.

4.2 Análise das possibilidades de agrupamento e seleção do agrupamento

A visualização das possíveis opções de agrupamento está representada na Figura 9. Posto isto, a Tabela 8 oferece uma síntese das matrizes que deram origem aos diferentes cenários de potenciais agrupamentos. Nela, estão contempladas as diversas hipóteses de agrupamentos, o número de agrupamentos efetivamente criados em cada uma das matrizes, bem como as duas métricas escolhidas para avaliar o desempenho do SOM, nomeadamente *quantization error* (QE) e *topographic error* (TE).

Tabela 8- Síntese dos vários cenários e cálculo das métricas de medição da performance

Matriz	Output (nº de agrupamentos obtidos)	Métricas
5x5	25/25 agrupamentos	QE = 6,49
		TE = 0,72
6x6	36/36 agrupamentos	QE = 6,44
		TE = 0,82
7x7	49/49 agrupamentos	QE = 6,41
		TE = 0,90
8x8	54/64 agrupamentos	QE = 6,25
		TE = 0,88
9x9	73/81 agrupamentos	QE = 6,32
		TE = 0,89
10x10	78/100 agrupamentos	QE = 6,30
		TE = 0,87
11x11	80/121 agrupamentos	QE = 6,26

		TE = 0,82
12x12	100/144 agrupamentos	QE = 6,24
		TE = 0,69
13x13	95/169 agrupamentos	QE = 6,21
		TE = 0,71
14x14	120/196 agrupamentos	QE = 6,18
		TE = 0,55
15x15	125/225 agrupamentos	QE = 6,16
		TE = 0,59

Fonte: Elaboração própria

Com base nos valores apresentados acima, e considerando o que foi discutido em 3.2.5, ao expor ambas as métricas, pode-se concluir que quanto menor o valor dessas métricas, melhor é a qualidade do mapa e maior é a preservação da sua topologia, respetivamente.

Embora o algoritmo tenha sido testado até $n=15$, a partir da matriz 14x14, a visualização dos mapas e as duas métricas de avaliação tornam-se semelhantes, à medida que n aumenta. Por essa razão, a matriz 14x14 destaca-se, de uma forma geral, pois apresenta valores mais baixos nestas duas métricas de avaliação do modelo, com $QE = 6,18$ e $TE = 0,55$.

Após uma análise estatística preliminar dos dados recolhidos, representada pelas medidas de tendência central (média e desvio-padrão), segue-se para a visualização das matrizes formadas, com base nos parâmetros definidos em 3.2.4.

A Figura 9 demonstra as seis matrizes que deram origem aos vários cenários de agrupamentos das observações em estudo, a iniciar na matriz 10x10, visto que é a partir desta que não se verifica grandes diferenças face às restantes, e as métricas de avaliação do SOM tornaram-se mais satisfatórias. Como mencionado anteriormente, foram calculadas duas métricas para avaliar a performance dos diferentes mapas, as quais também estão representadas na mesma figura, abaixo de cada um dos mapas.

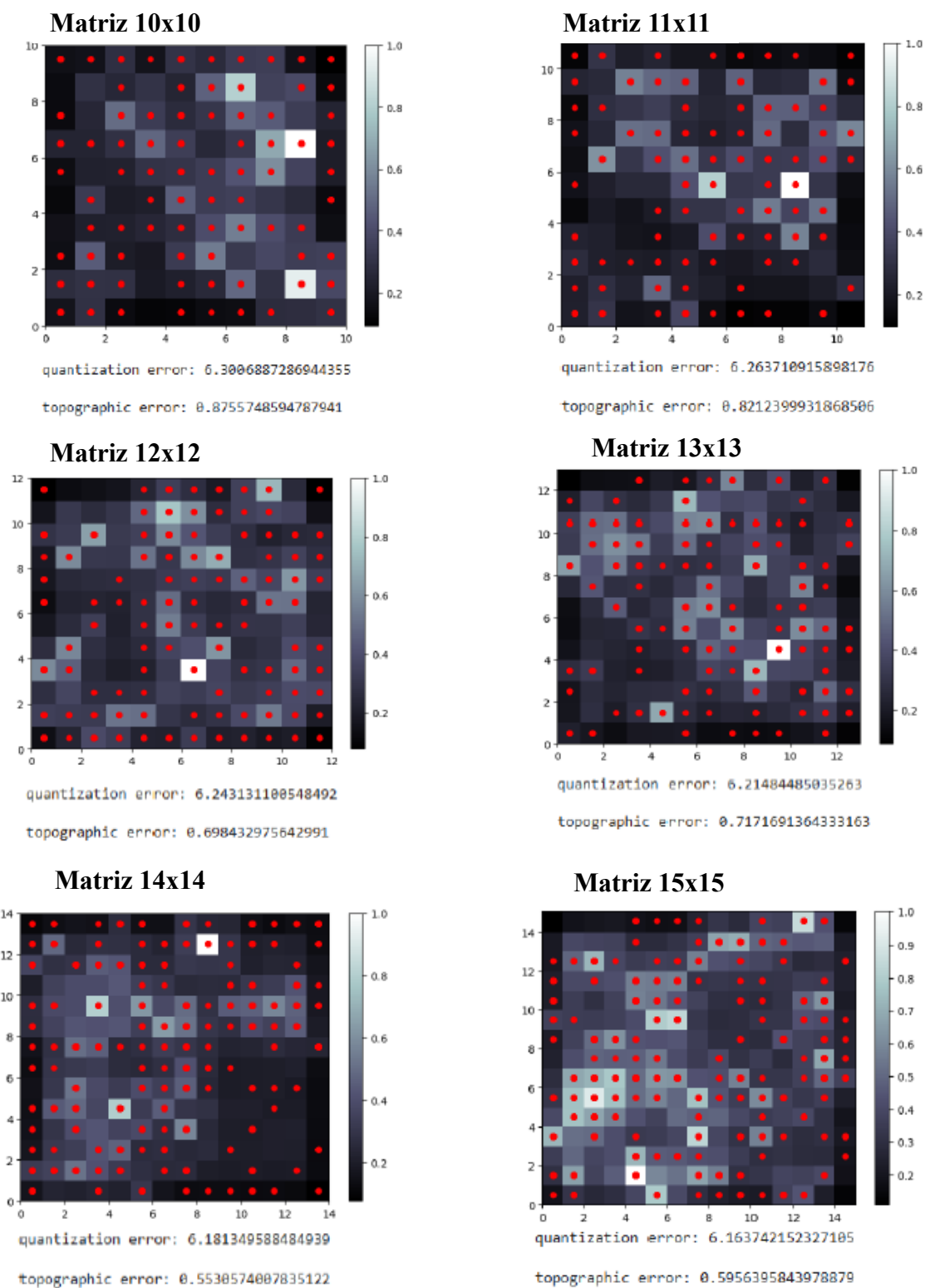


Figura 9- Demonstração das matrizes de agrupamento em Python

4.3 Determinação do nível de desempenho

A matriz de dimensão 14x14 e a matriz a partir da qual foram analisadas as escolas em questão, de acordo o *output* formado pelo código (MiniSom) e posterior caracterização dos seus agrupamentos, representada pela Figura 10. Para o tema em análise – sucesso escolar – foram definidos os seguintes níveis de desempenho: Excelente, Bom, Mau, Muito Mau. Estes níveis foram determinados com o objetivo de categorizar o desempenho das escolas de forma clara e intuitiva, facilitando tanto a análise quanto a comparação entre diferentes grupos.

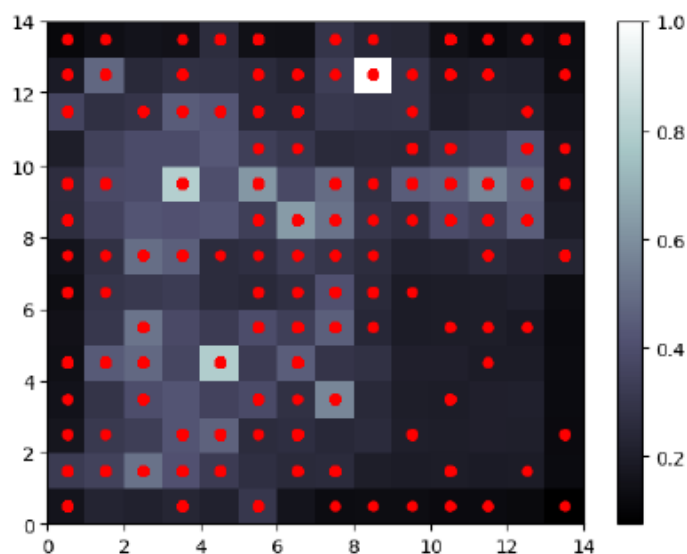


Figura 10- Ilustração da matriz escolhida em Python

Com base na definição da BMU e na análise de vizinhança, identificaram-se quatro classes, associadas a cada uma das escalas de cor do mapa SOM selecionado.

- Para os agrupamentos formados na matriz 14x14 e a partir do cálculo da média das 33 variáveis de avaliação do sucesso escolar para cada agrupamento, obteve-se a primeira classe, associada à escala preta, agrupou 6 observações que correspondem à média mais elevada das 30 variáveis de avaliação do sucesso escolar. Esta classe reúne as observações que são classificadas como excelentes;
- A segunda classe, relacionada com a escala cinzenta, agrupou 68 observações, que corresponde ao nível de desempenho bom;
- A terceira classe relacionada com a escala azul, agrupou 44 observações, que corresponde ao nível de desempenho mau;
- A quarta classe relacionada com a escala branca, agrupou 2 observações, que corresponde ao nível de desempenho muito mau.

4.4 Posicionamento das escolas face ao sucesso escolar

Para compreender as variáveis que mais influenciam o sucesso escolar, foi criado um gráfico de correlação em Python (*vide* Anexo C) que analisa a relação entre todas as variáveis do modelo. A partir dessa análise, destacam-se algumas variáveis que apresentaram uma correlação mais forte entre si: a percentagem de pais estrangeiros, a percentagem de alunos estrangeiros, a taxa de desistência, a taxa de retenção, a média das notas dos exames e provas nacionais, e a percentagem de alunos no 2º e 3º ciclos.

Estas variáveis foram selecionadas pois exibem uma relação direta com elementos que podem afetar o desempenho escolar dos alunos. A elevada percentagem de pais e de alunos estrangeiros, por exemplo, pode estar associada a desafios específicos, como barreiras linguísticas, diferenças culturais e dificuldades de adaptação ao sistema de ensino. Esses fatores podem influenciar a integração dos alunos e, por consequência, o seu desempenho académico.

Já a taxa de desistência e de retenção são indicadores fundamentais do progresso escolar. A desistência reflete uma rutura no percurso académico, o que impacta negativamente o sucesso escolar. A retenção, por outro lado, indica dificuldades de aprendizagem e falha no acompanhamento do currículo escolar, o que resulta numa menor probabilidade de sucesso escolar.

A média das notas em exames e provas nacionais oferece uma medida objetiva e padronizada do desempenho académico, pois permite avaliar o nível de aprendizagem em comparação com outros alunos. Esta variável é crucial, porque representa o conhecimento adquirido ao longo do percurso escolar e é um indicador direto de sucesso escolar.

Finalmente, a percentagem de alunos nos 2º e 3º ciclos mostra a distribuição dos estudantes. A concentração de alunos em determinados ciclos não é, por si só, uma variável que dificulte a progressão no ensino, mas pode estar associada a fatores como os anos que compõem cada ciclo (dois anos no 2º ciclo e três anos nos 1º e 3º ciclos) ou às taxas de retenção, que são mais elevadas em anos específicos, como o 5º, 7º e 9º anos. Assim, a não progressão dos alunos pode contribuir para a concentração em determinados ciclos ou anos de escolaridade.

Na sua generalidade, tendo em conta as variáveis mais correlacionadas, é possível compreender que, à semelhança do que anteriormente foi mencionado, fatores estruturais, culturais e pedagógicos estão diretamente associados ao (in)sucesso dos alunos.

Além de analisar as variáveis que mais influenciam o sucesso escolar através do gráfico de correlação (*vide* Anexo C), foi ainda criado um SOM com o intuito de agrupar as regiões com base nas semelhanças entre as escolas e o sucesso escolar. A Figura 11 ilustra a matriz onde é possível identificar os padrões e agrupamentos.

Primeiramente, importa realçar que o fundo do gráfico, em diferentes tons de cinza, representa o mapa de distância do SOM, i.e., as áreas claras ou brancas indicam regiões onde os neurónios do SOM estão mais distantes uns dos outros, sugerindo uma menor similaridade entre os perfis ali representados, por outro lado, as áreas escuras indicam neurónios mais próximos, sugerindo uma maior semelhança entre os dados. Estas áreas tendem a formar *clusters* mais definidos.

Posto isto, torna-se pertinente realizar uma análise mais detalhada às diferentes regiões representadas na matriz.

As escolas localizadas na região do Alentejo (cor vermelha) apresentam uma ampla dispersão no mapa, com pontos vermelhos difusos por várias áreas, tanto claras como escuras. Esta situação indica que as escolas do Alentejo apresentam uma ampla diversidade nos níveis de sucesso escolar.

Por outro lado, as escolas localizadas na região do Algarve (cor verde) mostram uma elevada concentração dos pontos numa área específica do mapa. Esses pontos estão posicionados numa região de fundo escura, pelo que as escolas do Algarve têm um perfil homogéneo nos níveis de sucesso escolar.

Já as escolas localizadas na região do Centro (cor azul) estão representadas em *clusters* nos limites superiores e inferiores do mapa. Esta análise sugere que existem grupos de escolas do Centro com perfis de sucesso escolar semelhante.

Na região de Lisboa e Vale do Tejo é onde se verifica um nível mais consistente e elevado de sucesso escolar. As escolas estão concentradas numa área escura e agrupada do mapa (cor roxa).

Por fim, as escolas localizadas na região Norte apresentam uma ampla variedade nos níveis de sucesso escolar, visto que os seus pontos se situam ao longo das extremidades da matriz.

A análise do SOM revela que os níveis de sucesso escolar variam significativamente entre regiões. Enquanto o Alentejo e o Norte exibem resultados que sugerem que as escolas dessas regiões enfrentam condições variáveis que afetam o sucesso escolar, Lisboa e Vale do Tejo e Algarve apresentam um perfil mais homogéneo, o que possivelmente indica um nível de sucesso escolar mais equilibrado.

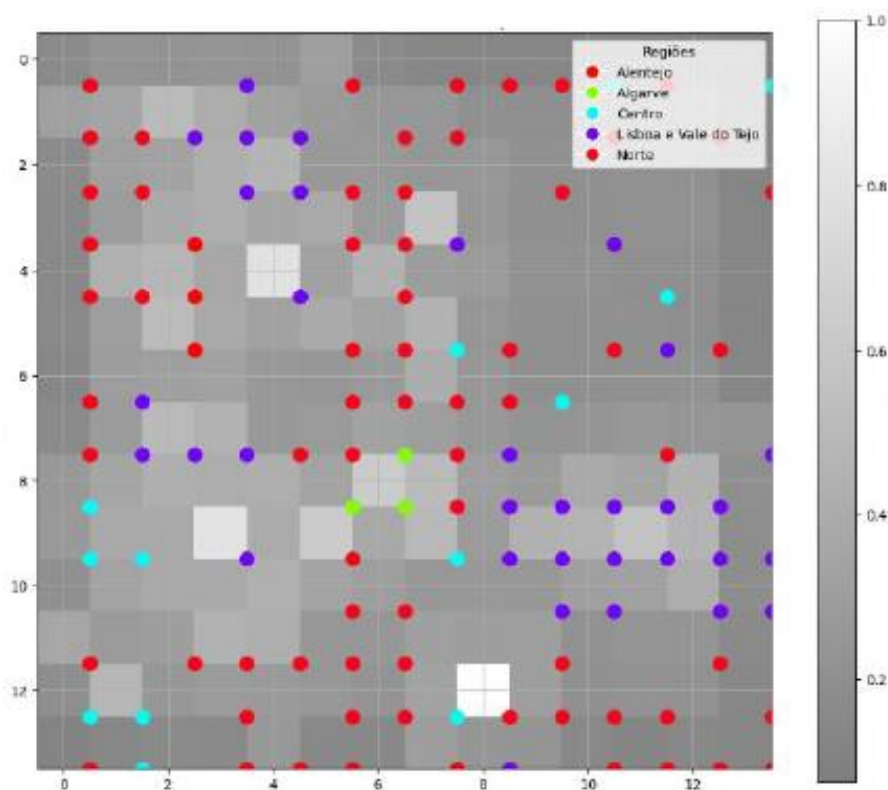


Figura 11- Ilustração da matriz escolhida em Python

4.5 Análise das questões de investigação

Na fase inicial da dissertação, foram formuladas questões de investigação que, após a aplicação da metodologia e a análise dos dados, agora podem ser avaliadas.

A primeira questão está associada aos indicadores e variáveis que contribuem para o sucesso escolar no nível básico do ensino público. Conforme a análise realizada, o sucesso escolar deve considerar, no mínimo, três dimensões: a dimensão socioeconómica, a dimensão organizacional e os indicadores de sucesso escolar, seis indicadores e trinta e duas variáveis, com base na fonte de dados utilizada.

A segunda questão pretendia refletir sobre as áreas geográficas onde se regista um maior sucesso escolar. Como referido anteriormente, concluiu-se que as regiões de Lisboa e Vale do Tejo e Algarve são aquelas onde existe um maior índice de sucesso escolar. Pelo contrário o Norte e o Alentejo é onde se verifica uma maior dispersão.

Realço que, embora a potencialidade do SOM esteja relacionada com a análise de dados a um nível mais granular, como por exemplo o de escolas, a análise por áreas geográficas oferece uma compreensão mais profunda dos fatores que influenciam o sucesso escolar num contexto

mais amplo. A opção de agrupar os dados por região, permitiu identificar padrões e tendências que não seriam visíveis numa análise por escolas. Por exemplo, regiões com maiores índices de sucesso escolar podem estar a beneficiar de melhores infraestruturas ou políticas públicas mais eficazes. Esta abordagem permite um diagnóstico mais holístico, que tem em consideração o contexto de cada área geográfica, o que facilita a identificação de causas subjacentes ao sucesso escolar. Além disso, ao trabalhar com dados por regiões, é possível observar escolas que compartilhem características semelhantes, como níveis socioeconómicos, permitindo uma análise mais contextualizada.

Por último, a terceira questão estava relacionada com a eficiência do SOM como ferramenta de análise do sucesso escolar. Tendo por base a sistematização realizada identificou-se uma variedade de modelos capazes de avaliar o sucesso escolar. Entre eles, destacam-se os modelos de aprendizagem supervisionada, como a classificação e a regressão, os modelos de aprendizagem não supervisionada, como o *K-means* e o SOM, e os modelos descritivos, incluem técnicas de agrupamento, clusterização e associação. O SOM foi o modelo escolhido para aplicar devido às suas vantagens, entre as quais se destacam:

- A capacidade de representar dados complexos, com muitas variáveis, num espaço bidimensional, facilitando a visualização e a identificação de padrões nos dados analisados;
- A formação de agrupamentos que podem ser interpretados intuitivamente, pois o SOM organiza os dados de forma natural, através da criação de agrupamentos de neurónios que correspondem a áreas similares no espaço de origem;
- A sua ampla aplicabilidade em diversos domínios, como a análise de dados geoespaciais, processamento de imagens, reconhecimento de padrões, entre outros exemplos;
- A simplicidade no treinamento e na interpretação dos mapas, em comparação com redes neurais mais complexas, que requerem abordagens mais elaboradas para o seu uso e compreensão.

A escolha do SOM em detrimento de outros modelos não supervisionados, como por exemplo o *K-means*, está relacionada ao facto de que para este último, é vantajoso ter uma ideia inicial sobre o problema e os resultados esperados. Nesse contexto, o *K-means* é mais adequado para análises exploratórias de dados, enquanto neste estudo, o número de agrupamentos que iriam ser formados eram desconhecidos desde o início.

Capítulo 5. Conclusões

5.1 Recomendações da investigação

Esta dissertação demonstrou que o SOM oferece benefícios significativos para a análise do sucesso escolar. Uma das principais recomendações é a implementação de grelhas de avaliação padronizadas e fixas nas escolas, permitindo que o processo iterativo de aprendizagem do SOM seja comparável ao longo dos diferentes momentos de avaliação. Outra recomendação decorrente deste estudo está relacionada com a monitorização contínua dos indicadores de sucesso escolar, com integração de dados de múltiplos anos letivos, para otimizar a identificação de padrões e tendências ao longo do tempo.

Esta dissertação também enriquece o conhecimento científico e académico, além de oferecer contribuições relevantes para os profissionais da área da educação, na medida em que fornece uma abordagem inovadora de análise de dados através do SOM, que permite identificar padrões de sucesso e desafios no desempenho escolar. Esta contribuição revela-se pertinente, pois parece evidente que as escolas não têm demonstrado a capacidade de implementar estratégias inclusivas para alunos imigrantes, o que se reflete em maiores índices de insucesso escolar nas escolas com maior percentagem destes alunos. Ou seja, o contributo visa evidenciar essa lacuna, o que pode incitar um processo reflexivo por parte dos profissionais da educação, permitindo a melhoria dos mecanismos internos de correção e a adoção de abordagens mais eficazes na promoção da inclusão. Desta forma é possível promover uma melhoria contínua na qualidade do ensino e na eficácia das intervenções educativas.

5.2 Limitações da investigação

Quanto às limitações da investigação, foram identificados diversos desafios ao longo da elaboração desta dissertação. O primeiro obstáculo está relacionado com a variabilidade nas práticas de avaliação entre escolas. Mesmo com grelhas de avaliação constantes, práticas pedagógicas e de avaliação podem variar entre escolas e professores, o que pode introduzir ruído nos dados e dificultar a comparação direta entre diferentes contextos escolares.

Outra das limitações está relacionada ao facto do estudo estar limitado ao ensino básico público. Ao centrar-se apenas nas escolas públicas do ensino básico, os resultados não são generalizáveis para o ensino secundário ou para o setor privado, limitando a aplicabilidade da análise a contextos educativos diferentes.

Além disso, existem falhas de informação relativas aos indicadores socioeconómicos, que decorrem, por vezes, do não preenchimento ou do preenchimento incorreto desses dados, seja

por erros cometidos pelas famílias, seja pelas próprias escolas. Esta lacuna pode comprometer a precisão das análises e limitar a validade dos resultados obtidos.

Para terminar, o uso do SOM exige conhecimentos técnicos avançados em análise de dados e acesso a recursos computacionais adequados. No decorrer da revisão literária, é visível a escassez de recursos sobre esta temática em países de língua oficial portuguesa. A maior parte dos estudos relacionados com este modelo e técnica encontra-se em âmbito internacional. Essa realidade é evidenciada no capítulo seguinte, que apresenta as referências bibliográficas utilizadas para a realização desta dissertação.

5.3. Sugestões para investigações futuras

Entre as principais sugestões para investigações futuras destaca-se a possibilidade de incorporar um maior número de variáveis no algoritmo do SOM. Por exemplo, integrar novas variáveis além das 33 utilizadas de forma a diversificar o leque de informação. Torna-se interessante diversificar as variáveis da dimensão organizacional.

Outra sugestão prende-se com a realização de estudos longitudinais, integrando mais de três anos letivos, que acompanhem o progresso dos estudantes ao longo do tempo para identificar padrões de sucesso escolar ou dificuldades de aprendizagem. Este tipo de análise iria ajudar a compreender o impacto de mudanças educativas e contextuais ao longo dos anos.

Expandir a aplicação do SOM a outros níveis de ensino, como o secundário e o superior, bem como também, ao setor privado. Esta comparação poderia gerar *insights* relevantes sobre práticas pedagógicas e políticas educativas em diferentes contextos.

Com complemento à análise efetuada, seria relevante combinar o SOM com outras técnicas de análise de dados. Esta abordagem permitiria comparar os resultados e avaliar o tipo de agrupamento gerado pelo modelo, contribuindo para verificar e até validar se as escolas foram agrupadas de forma adequada e conforme as semelhanças observadas em relação ao sucesso escolar.

Referências Bibliográficas

- Afonso, A. J. (2009). Nem tudo o que conta em educação é mensurável ou comparável. Crítica à accountability baseada em testes estandardizados e rankings escolares. *Revista Lusófona de Educação*, 13, pp. 13-29.
- Ali, A. & Amin, M., Z. (2019). An Intuitive Guide of Self Organizing Maps with Practical Implementation in Minisom. Wavy AI Research Foundation. Disponível em https://www.academia.edu/41216265/An_Intuitive_Guide_of_Self_Organizing_Maps_with_Practical_Implementation_in_Minisom.
- Almada, A. (2015). Políticas sociais de educação no sucesso educativo. *Diversidades: Sucesso Escolar*, 47, pp. 15-16.
- Álvares, M., Calado, A. (2014). Insucesso e Abandono Escolar Precoce – Os programas de apoio, 40 Anos de Políticas de Educação em Portugal (Vol. I, pp-1-25). Almedina. https://www.researchgate.net/publication/296648198_Insucesso_e_Abandono_Escolar_Precoces_-_Os_programas_de_apoio
- Alves, C. (2010). Insucesso escolar em Língua Portuguesa - um estudo caso. Lisboa: Escola Superior de Educação João de Deus.
- Alves, J. M. & Cabral, I. (2016a). Condições políticas, organizacionais e profissionais da promoção do sucesso escolar – ensaio de síntese. In Formosinho, J. et al. (org.), *Nova organização da escola. Caminhos de possibilidades*, Fundação Manuel Leão, Vila Nova de Gaia, pp. 19-38.
- Amorim, S. M. S. (2019). Fatores Organizacionais e (in)sucesso escolar: um estudo de caso duplo comparativo [Dissertação de Doutoramento em Ciências da Educação, Universidade Católica Portuguesa]. Repositório da Universidade Católica Portuguesa. [Veritati - Repositório Institucional da Universidade Católica Portuguesa: Fatores organizacionais e \(in\)sucesso escolar : um estudo de caso duplo comparativo \(ucp.pt\)](#)
- Asan, U. & Ercan, S. (2012). An Introduction to Self-Organizing Maps. Em C. Kahraman. *Computational Intelligence Systems in Industrial Engineering: With Recent Theory and Applications*, (pp. 299-319). Atlantis Press. Disponível em DOI:10.2991/978-94-91216-77-0_14.
- Benavente, A. (1990). Insucesso escolar no contexto português: abordagens, concepções e políticas, *Análise Social*, XXV (108-109), pp. 715-733.
- Branco, P. (2012). Insucesso escolar: um estudo na área escolar da Maia. Porto: Universidade Fernando Pessoa.

- Cabral, I. (2014). Gramática Escolar e (in)sucesso - Os Projetos Fénix, Turma Mais e ADI. Porto: Universidade Católica Portuguesa.
- Carmo, R. M. (Coord.) (2024). *Jovens e o Trabalho em Portugal. Desigualdades, (Des)Proteção e Futuro*, Edições Almedina.
- Claro, C. (2015). Insucesso e Sucesso na Organização Escolar. *Organização pedagógica e diversidade*. Iscte- Instituto Universitário de Lisboa.
- CNE - Conselho Nacional de Educação (2017). Perfil do Aluno: Competências para o século XXI, Relatório Técnico. [Relatório Técnico: Perfil do Aluno: Competências para o século XXI - Conselho Nacional de Educação \(cneu.pt\)](https://cneu.pt/RelatorioTecnico/Perfil-do-Aluno-Competencias-para-o-seculo-XXI)
- Decreto-Lei n.º 46/86. (1986). Diário da República, 1.ª Série – nº 237.
- Decreto-Lei n.º 75/08. (2008). Diário da República, 1.ª Série – nº 79.
- Decreto-Lei n.º 176/12. (2012). Diário da República, 1.ª Série – nº 149.
- Figueiredo, C. (1975). Pequeno dicionário da língua portuguesa. Amadora: Bertrand. Flick, U. (2009). Desenho da pesquisa qualitativa. Porto Alegre: Artmed.
- Franco, L. M., Bennett, S., & Kanfer, R. (2002). Health sector reform and public sector health worker motivation: a conceptual framework. Disponível em: https://www.researchgate.net/publication/11380270_Public_Sector_Health_Worker_Motivation_and_Health_Sector_Reform_A_Conceptual_Framework Acessado 30 de janeiro.
- Lei n.º 7-B/2016, de 31 de março de 2016. http://bdjur.almedina.net/item.php?field=node_id&value=2101747
- Forest, F., Lebbah, M., Azzag, H. & Lacaille, J. (2020). A Survey and Implementation of Performance Metrics for Self-Organized Maps. *arXiv*. Disponível em <https://doi.org/10.48550/arXiv.2011.05847>.
- Grimheden, S. & Jarlesater, J. (2022). *Concretizing CRISP-DM for Data-Driven Financial Decision Support Tools*. [Projeto de mestrado, LTH - Lund University].
- Kalteh, A. M., Hjorth, P. & Berndtsson, R. (2008). Review of the Self-Organizing Map (SOM) approach in water resources: analysis, modelling and application. *Environmental Modelling & Software*, 23(7), 835-845. Disponível em DOI:10.1016/j.envsoft.2007.10.001.
- Langone, R., Alzate, C., De Ketelaere, B., Vlasselaer, J., Meert, W. & Suykens, J. (2015). LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines. *Engineering Applications of Artificial Intelligence*, 37(5), 268-278. Disponível em <https://doi.org/10.1016/j.engappai.2014.09.008>.
- Lemos, V. (2013). *Políticas públicas de educação: equidade e sucesso escolar*. Sociologia, Problemas e Práticas, 73, 151-169.

- Lemos, V. V. (2014). A influência da OCDE nas políticas públicas de educação em Portugal [Tese de Doutoramento em Políticas Públicas]. Repositório do Iscte. [Repositorio do Iscte – Instituto Universitário de Lisboa: A influência da OCDE nas políticas públicas de educação em Portugal \(iscte-iul.pt\)](https://repositorio.iscte-iul.pt/handle/10071/10071)
- Martins, H. G. (2010). *Causas de (in)sucesso escolar* [Dissertação de Mestrado, Universidade de Trás-os-Montes e Alto Douro]. Repositório UTAD.
- Martins, H. (2017). *Insucesso Escolar Prevenção e Intervenção na Educação Pré-Escolar e no 1º Ciclo do Ensino Básico*. [Dissertação de Mestrado, Escola Superior de Educação de Paula Frassinetti]. Repositório da Escola Superior de Educação de Paula Frassinetti. <http://repositorio.esepf.pt/bitstream/20.500.11796/2504/1/FINAL.pdf>
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). Instrução em sala de aula que funciona: estratégias baseadas em pesquisa para aumentar o desempenho dos alunos. ASCD.
- Mendonça, A. (2006). A problemática do insucesso escolar. A escolaridade obrigatória no arquipélago da Madeira em finais do século XX (1994-2000) [Dissertação de Doutoramento em Sociologia da Educação, Universidade da Madeira]. Repositório da Universidade da Madeira. [DigitUMa: A problemática do insucesso escolar: a escolaridade obrigatória no Arquipélago da Madeira em finais do século XX \(1994-2000\)](https://digituma.uma.pt/bitstream/10216/10071/1/A%20problem%C3%A1tica%20do%20insucesso%20escolar%20a%20escolaridade%20obrigat%C3%B3ria%20no%20Arquip%C3%A9lago%20da%20Madeira%20em%20finais%20do%20s%C3%A9culo%20XX%20(1994-2000).pdf)
- Miljković, D. (2017, Maio). Brief Review of Self-Organizing Maps. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1061-1066. Disponível em DOI:10.23919/MIPRO.2017.7973581.
- Miranda, C. (2010). *Causas de (in)Sucesso escolar Escolar*. [Dissertação de Mestrado, Universidade de Trás-os-Montes e Alto Douro].
- Nagashima, H. & Kato, Y. (2019). APREP-DM: a Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM. 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japão, 555-560, Disponível em DOI: 10.1109/PERCOMW.2019.8730785.
- Narciso, A. R. L. (2023). Avaliação do comportamento das agências bancárias numa perspetiva de qualidade do serviço [Dissertação de Mestrado, ISCTE-Instituto Universitário de Lisboa]. Repositório ISCTE-IUL. [master_ana_lincho_narciso.pdf](https://repositorio.iscte-iul.pt/handle/10071/10071)
- Negri, R. G. (2013). At-SOM: Mapas Auto-organizáveis atenuantes. *Revista Brasileira de Cartografia*, 1-9. https://www.researchgate.net/publication/253328440_At-SOM_MAPAS_AUTO-ORGANIZAVEIS_ATENUANTES

- Nunes, L.C. (Coord.) (2023). *Da desigualdade social à desigualdade escolar nos municípios de Portugal*, Fundação Belmiro de Azevedo.
- Pedroso, V. (2022). *Conjunto de materiais: educação inclusiva*. <https://www.dge.mec.pt/sites/default/files/ebooks/gestao-da-educacao-inclusiva.pdf>
- Peixoto, M. (1999). Auto-estima, inteligência e sucesso escolar : um estudo com alunos do 9º ano de escolaridade. Braga: APPACDM.
- Rangel, A. (1994). Insucesso escolar . Lisboa: Instituto Piaget.
- Ribeiro, I. S., Almeida, L. S., Gomes, C. (2006). Conhecimentos prévios, sucesso escolar e trajetórias de aprendizagem: do 1º para o 2º ciclo do ensino básico. Avaliação Psicológica, 5(2), 127-133. <http://www.redalyc.org/articulo.oa?id=335027180002>
- Roazzi, A., Almeida, L. S. (1988). Insucesso escolar: insucesso do aluno ou insucesso do sistema escolar?. Revista Portuguesa de Educação, 1(2), 53-60. https://www.researchgate.net/publication/277146171_Insucesso_escolar_insucesso_do_aluno_ou_insucesso_do_sistema_escolar
- Rosa, M. (2013). Causas de Abandono e Insucesso Escolar. Vila Real: Universidade de Trás-Os-Montes e Alto Douro.
- Silva, S., & Fraga, N. (2021). Autonomia e Flexibilidade Curricular como Instrumentos Gestionários. O Caso de Portugal. *REICE. Revista Iberoamericana Sobre Calidad, Eficacia Y Cambio En Educación*, 19(2), 37–54.
- Soares, D. R. L. (2015). *Sucesso escolar e percursos vocacionais na adolescência: variáveis pessoais e contextuais* [Tese de Doutoramento, Universidade do Minho]. Repositorium SDUM. [Diana Rafaela Lopes Soares pos defesa](#)
- Utsch, A., Morchen, F. (2005). *ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM*. Data Bionics Research Group, University of Marburg, Germany.
- Vasconcelos, M. (2015). O Insucesso Escolar. *Diversidades: Sucesso Escolar*, 47, pp. 6-9.
- Ventura, A. (2021). Repensar o sucesso escolar: Para uma abordagem mais inclusiva e equitativa, *Revista Portuguesa de Educação*, 34(2), 45-62.
- Ventura, A., Castanheira, P., & Adelino Costa, J. (2006). Gestão das escolas em Portugal. *REICE: Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 4(4), 128-136.
- Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). SOM Toolbox for Matlab 5. Disponível em: <http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf>
Acessado em 28 de janeiro.

Vieira, M. (2013). *Gramática escolar e (in)sucesso – os casos do projeto Fénix, Turma Mais e ADI* [Tese de Doutoramento, Universidade Católica Portuguesa]. Repositório da Universidade Católica Portuguesa.
[https://repositorio.ucp.pt/bitstream/10400.14/12584/1/Disserta%
c3%a7%c3%a3o_Il%
c3%addia_Vieira_15-04-2013.pdf](https://repositorio.ucp.pt/bitstream/10400.14/12584/1/Disserta%c3%a7%c3%a3o_Il%c3%addia_Vieira_15-04-2013.pdf)

Anexos

Anexo A- Grelha de avaliação

Variáveis ⁵	Fonte	Definição
Indicadores socioeconómicos dos alunos		
Percentagem de alunos com ASE (escalão A e B)	DGEEC	Conjunto de medidas e apoios financeiros, diretos e indiretos, concedidos pelo Estado a fundo perdido, que visam garantir o direito e igualdade de oportunidades de acesso à educação, frequência e sucesso escolares, através da superação de desigualdades económicas, sociais e culturais
Percentagem de alunos com PEI	DGEEC	Documento que fixa e fundamenta as necessidades especiais da criança ou do jovem e respetivas formas de avaliação baseadas na observação e avaliação em sala de aula e nas informações complementares disponibilizadas pelos participantes no processo individual do aluno, segundo legislação em vigor
Percentagem de alunos com nacionalidade estrangeira	DGEEC	Os dados referem-se à percentagem de alunos matriculados no ensino básico que não possuem nacionalidade portuguesa
Percentagem de alunos com pais encarregados de educação com nacionalidade estrangeira	DGEEC	Os dados referem-se à percentagem de alunos matriculados no ensino básico cujos pais encarregados de educação não possuem nacionalidade portuguesa
Percentagem de encarregados de educação com ensino secundário ou superior completo	DGEEC	Os dados referem-se à percentagem de encarregados de educação que possuem o ensino secundário ou superior completo
Percentagem dos tipos de situação profissional dos pais ou encarregados de educação	DGEEC	Os dados referem-se à percentagem dos diferentes tipos de situação profissional dos pais ou encarregados de educação dos alunos

⁵ São apresentadas apenas as variáveis que carecem de definição

		matriculados no ensino Básico
Percentagem dos tipos de profissões dos pais ou encarregados de educação	DGEEC	Os dados referem-se à percentagem dos diferentes tipos de profissões dos pais ou encarregados de educação dos alunos matriculados no ensino Básico
Indicadores organizacionais		
Percentagem de docentes do quadro	DGEEC	Docente com contrato de trabalho em funções públicas por tempo indeterminado, tendo em vista a satisfação das necessidades dos jardins de infância ou dos estabelecimentos de ensino
Percentagem de docentes contratados	DGEEC	Docente com contrato de trabalho em funções públicas a termo resolutivo, certo ou incerto, tendo em vista a satisfação de necessidades residuais do sistema de educação e formação não colmatadas pelo pessoal docente dos quadros
Número de docentes no ensino Básico	DGEEC	Docente habilitado para a docência destes níveis de ensino, formados como especialistas numa determinada área nas universidades. Inclui ainda pessoal docente portador dos requisitos exigidos para o acesso à profissionalização em exercício ou que dela tenha sido dispensado
Percentagem de docentes por grupos etários	DGEEC	Os dados referem-se à percentagem de docentes por grupos etários que lecionam ensino Básico
Número de estabelecimentos escolares que lecionam ensino Básico	DGEEC	Os dados referem-se apenas aos estabelecimentos de ensino que lecionam ensino Básico
Localização das Escolas	DGEEC	Os dados referem-se apenas à localização das que lecionam o ensino Básico
Indicadores de sucesso escolar		
Média das provas nacionais de exame no ensino Básico	DGEEC	Os dados referem-se à média obtida pelos alunos nas provas nacionais de exame no ensino Básico

Taxa de retenção do ensino Básico	DGEEC	Situação que ocorre em consequência do aproveitamento sem êxito do aluno pelo não cumprimento dos requisitos previstos na legislação em vigor para a frequência no ano de escolaridade seguinte àquele em que se encontra
Taxa de desistência do ensino Básico	DGEEC	Situação que ocorre em consequência do abandono temporário de aluno ou formandos da frequência das atividades letivas de um curso, de um período de formação ou de uma ou mais disciplinas no decurso de um ano letivo
Percursos Diretos de Sucesso do ensino Básico	Infoescolas (fonte de dados DGEEC)	Mostra os alunos que concluíram o ensino Básico dentro do tempo esperado, ou seja, até nove anos após a sua ingresso
Média de retenção por aluno do ensino Básico	DGEEC	Situação que ocorre em consequência do aproveitamento sem êxito do aluno pelo não cumprimento dos requisitos previstos na legislação em vigor para a frequência no ano de escolaridade seguinte àquele em que se encontra

Anexo B- Código SOM em *Python*

```
from minisom import Minisom
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pylab import bone, pcolor, colorbar, plot, show
from sklearn.preprocessing import StandardScaler

# Configuração e treinamento do SOM
data=df.values
n=14
som = Minisom(x=n, y=n, input_len=len(data[0]), sigma=0.7,
learning_rate=0.07,neighborhood_function='gaussian', random_seed=0)
som.train_random(data = data, num_iteration = 3000)

# Visualização do mapa SOM
bone()
pcolor(som.distance_map().T)
colorbar()
for i, x in enumerate(data):
    w = som.winner(x) # Agora esta linha está corretamente indentada
    plot(w[0] + 0.5, w[1] + 0.5, 'o', markeredgecolor='r', markerfacecolor='r', markersize=4, markeredgewidth=2)
show()

# Cálculo e exibição dos erros
print("\n\nquantization error:",som.quantization_error(data),"\n")
print("topographic error:",som.topographic_error(data),"\n\n\n")

# Coordenadas dos neurônios vencedores
winner_coordinates = np.array([som.winner(x) for x in data])
df['cluster'] = [f"{coord[0]}-{coord[1]}" for coord in winner_coordinates]

# Agrupamento das informações
grouped_data = df.groupby('cluster').mean()

# Exibir agrupamentos e principais características
print("\nMédia das características por grupo de clusters no SOM:\n")
print(grouped_data)
```

Nota: Parâmetros preenchidos para n=14

Anexo C- Gráfico de correlação em Python

