

O papel da Big Data na análise de dados climáticos: previsão de eventos climáticos extremos

Daniela Valente Rei, Instituto Politécnico de Castelo Branco, Portugal, daniela.rei@ipcbcampus.pt

Diogo Fernandes Marçal, Instituto Politécnico de Castelo Branco, Portugal,
d.marcal@ipcbcampus.pt

Eurico Lopes, Instituto Politécnico de Castelo Branco, Portugal, eurico@ipcb.pt

Arminda G. Lopes, Instituto Politécnico de Castelo Branco, Portugal, aglopes@ipcb.pt

Luis M. Barata, Instituto Politécnico de Castelo Branco, Portugal, luis.barata@ipcb.pt

Resumo

A *Big Data* pode ser aplicada a diversas áreas da atividade humana sendo uma ferramenta muito útil na previsão de acontecimentos. Uma área em que a *Big Data* pode ser aplicada é a previsão de eventos climáticos extremos como tornados, secas ou ondas de calor que muito prejudicam a vida das populações.

Analisando grandes volumes de dados climáticos, é possível prever a ocorrência de catástrofes relacionadas com esses dados, de modo a minimizar os efeitos adversos que provocam em infraestruturas e pessoas. A principal motivação em muitas aplicações da *Big Data* a esta temática é salvar vidas, mas também munir agricultores ou empresários de ferramentas de previsão para agirem com antecedência de modo a não sofrerem tantos prejuízos.

Justifica-se assim a elaboração de uma revisão sistemática que faça um levantamento de artigos que abordam a utilização da *Big Data* na previsão de fenómenos climáticos extremos.

Palavras-chave: *Big Data*; Previsão; Fenómenos climáticos extremos; Dados climáticos

Abstract

Big Data can be applied to several areas of human activity being a very useful tool in predicting events. One area where Big Data can be applied is in the prediction of extreme weather events such as storms, droughts or heat waves that greatly damage the lives of populations.

By analyzing large volumes of climate data, it is possible to predict the occurrence of related disasters in order to minimize the adverse effects they cause on infrastructure and people. The main motivation in many applications of Big Data in this area is to save lives but also to provide farmers or entrepreneurs with forecasting tools to act in advance so as not to suffer so much damage.

This justifies the development of a systematic review that surveys articles that address the use of Big Data in predicting extreme weather events.

Keywords: *Big Data; prediction; extreme weather phenomena; climate data*

1. INTRODUÇÃO

Esta revisão sistemática aborda a utilização da *Big Data* na análise de dados climáticos com o objetivo de efetuar previsões sobre eventos climáticos extremos. Com este trabalho tem-se como

objetivo compreender a importância da previsão de fenómenos climáticos adversos através da análise de artigos científicos que abordam o tema em questão.

Catástrofes naturais causadas por fenómenos climáticos extremos como tornados, cheias, tempestades, secas ou ondas de calor têm um impacto devastador não só nas atividades económicas desenvolvidas pelo ser humano, mas também na qualidade de vida das pessoas. Catástrofes como as enunciadas acima, podem provocar estragos em infraestruturas provocando crises e podem ainda provocar a perda de vidas humanas, daí a importância que o tema abordado pela revisão sistemática assume para prever tais fenómenos a fim de pelo menos tentar minimizar os seus efeitos devastadores. No caso específico de um tornado, a título de exemplo, conseguindo prever a sua trajetória é possível evacuar regiões por onde se espera que o tornado passe evitando a perda de vidas.

Para melhor compreensão desta revisão sistemática, é conveniente perceber o conceito de *Big Data*. De forma muito breve, *Big Data* consiste em grandes volumes de dados de diversas fontes que possuem valor através de informações, padrões e *insights* que podem ser vantajosos conhecer para as entidades e organizações a quem esses dados interessam.

Em cada artigo analisado pretende-se identificar a resposta a questões como as áreas de atividade humana abrangidas pela previsão de eventos climáticos extremos, a fonte dos dados climáticos, qual o fenómeno extremo que se pretende prever e ainda se foi elaborado um sistema de previsão e se sim descrevê-lo.

O presente trabalho encontra-se segmentado em cinco capítulos sendo eles: **Introdução** onde se contextualiza o tema do trabalho e qual a sua importância; **Métodos**, onde se explicará o processo de seleção de artigos a analisar segundo vários critérios; **Resultados**, onde serão apresentadas as análises feitas a cada trabalho selecionado; **Discussão** em que está presente uma perspetiva geral das informações identificadas nos trabalhos analisados; e a **Conclusão** do presente trabalho.

2. MÉTODOS

Neste capítulo será apresentado o método/procedimento utilizado para que fosse possível chegar até aos trabalhos relevantes para este artigo indicando a fonte dos trabalhos relacionados e a pesquisa e filtragem efetuadas para obter apenas os trabalhos relacionados com a temática em questão. Antes de ser possível analisar os artigos relevantes foi necessário selecioná-los, uma vez que após uma pesquisa inicial nem todos os trabalhos encontrados estavam relacionados com a matéria que se pretende analisar. Esse processo de seleção é descrito de seguida.

Como fonte de pesquisa usou-se a biblioteca digital *IEEE Xplore* que contém artigos e livros científicos e técnicos apresentados pela IEEE (*Institute of Electrical and Electronics Engineers*). A biblioteca digital *IEEE Xplore* permite que um utilizador efetue uma pesquisa avançada através da

aplicação de filtros que utilizam operadores lógicos (por exemplo AND ou OR). O utilizador indica por que expressões pretende pesquisar aplicando essas expressões a diferentes campos dos documentos da biblioteca digital tais como o “*Document Title*”, “*Abstract*”, entre outros.

Começou-se por definir um conjunto de palavras/expressões pelas quais se orientaria a pesquisa e que seriam contempladas na “*query*” ou mais corretamente no “*Command Search*”. Formou-se então um conjunto de combinações sendo que maioria delas continha o termo “*Big data*”.

A pesquisa teve o seguinte “*Command Search*”.

```
("Abstract": "Big data" AND "Abstract": "weather prediction")  
OR ("Abstract": "Big data" AND "Abstract": "climate analysis")  
OR ("Abstract": "Big data" AND "Abstract": "weather events")  
OR ("Abstract": "Big data" AND "Abstract": "climate data")  
OR ("Abstract": "Big data" AND "Abstract": "extreme weather")  
OR ("Abstract": "Big data" AND "Abstract": "weather forecasting")  
OR ("Abstract": "weather damage*" AND "Abstract": "natural calamities")  
OR ("Abstract": "weather damage*" AND "Abstract": "Big data")  
OR ("Abstract": "Big data" AND "Abstract": "extreme climate")  
OR ("Abstract": "Big data" AND "Abstract": "extreme" AND "Abstract": "flooding" )  
OR ("Abstract": "Big data" AND "Abstract": "climate" AND "Abstract": "forecast" )  
OR ("Abstract": "hurricane forecasting" AND "Abstract": "meteorological" )
```

Definida a expressão que foi aplicada ao *abstract* dos trabalhos, esta foi usada no *IEEE Xplore* obtendo-se um total de 85 trabalhos. Em seguida definiu-se que apenas se pretendia obter os trabalhos cuja data de publicação fosse considerada recente pelo que se limitou a pesquisa de 2014 a 2023 tendo sido eliminado apenas um trabalho ficando a restar 84 trabalhos.

Posteriormente procedeu-se à leitura dos títulos e *abstracts* dos trabalhos publicados tendo-se eliminado um total de 76 trabalhos restando 8 para leitura.

Como critérios de inclusão definiu-se que os trabalhos a seleccionar deveriam abordar o conceito da *Big data* associado à análise de dados meteorológicos e ainda trabalhos publicados que abordem a importância dessa mesma análise à previsão de fenómenos climáticos extremos.

Após leitura minuciosa excluíram-se 3 trabalhos restando no final de todas estas fases um total de 5 trabalhos.

Chegados ao conjunto de trabalhos finais, estes serão analisados segundo alguns critérios em seguida enunciados tendo em conta que nem todos os artigos podem conter a resposta direta a estas questões:

- Quais a(o)s técnicas/métodos de *Big data* utilizados?
- Qual a origem dos dados (ao nível da fonte e/ou geograficamente)?
- Quais os principais problemas identificados?
- Qual a aplicação prática da utilização da *Big data* na previsão de fenómenos climáticos extremos e o que este processo pode melhorar na vida das sociedades?
- Quais as previsões para a área no futuro?
- Quais as conclusões dos trabalhos?

3. RESULTADOS

Nesta secção apresentam-se as análises feitas aos artigos e trabalhos que resultaram do método de pesquisa explicado no capítulo anterior e apresentados na tabela seguinte:

Autores	Ano	Título
Jain & Jain	2017	Big Data in Weather forecasting: Applications and Challenges
Cox et al.	2018	An Accurate Model for Hurricane Trajectory Prediction
Madan et al.	2018	Analysis of Weather Prediction using Machine Learning & Big Data
Khan et al	2020	Prediction of Flood Severity Level via Processing IoT Sensor Data Using a Data Science Approach
Li et al.	2020	Research and Design of Meteorological Disaster Early Warning System for Power Grid Based on Big Data Technology

Tabela 1 – Artigos incluídos na Revisão Sistemática

O trabalho publicado por Jain & Jain (2017), aborda a contribuição da *Big data* para melhorar a previsão meteorológica através da análise de dados sobre o clima. Desta forma é possível gerar informações úteis com antecedência para a previsão de fenómenos climáticos que possam ameaçar a segurança das sociedades e do funcionamento produtivo de diversas áreas de atividade económica. O trabalho em questão foca-se nas áreas de atividade em que a *Big data* aplicada à previsão meteorológica pode dar importantes informações, mas também nos desafios que a *Big data* enfrenta na área da previsão meteorológica.

Na área da previsão meteorológica, a *Big data* contribui nos nossos dias de forma notória permitindo antecipar eventos climáticos extremos antes que eles aconteçam dando a oportunidade a quem por eles é afetado de se preparar de forma a minimizar o impacto devastador que estes possam ter em várias áreas da atividade humana. Desta forma é possível não só a salvar vidas reduzindo os riscos de fenómenos perigosos, mas também aumentar a produtividade das pessoas e organizações no nível económico.

Antes da *Big data*, a previsão meteorológica era feita por pessoal especializado, mas com o advento da *Big data*, dados como a humidade, temperatura e a velocidade do vento são usados para esta área tendo a velocidade do processo de previsão aumentado significativamente. No trabalho aqui analisado é referido que cerca de 33% do PIB mundial é diretamente afetado pela meteorologia, facto revelador da importância da utilização da *Big data* na previsão meteorológica. Neste trabalho é referido que desde que a *Big data* é utilizada na previsão meteorológica a precisão/exatidão desta aumentou, bem como o número de pessoas que usam essas previsões na sua vida.

No que toca ao turismo, atividade com uma importância económica crescente, a prevenção de fenómenos climáticos adversos é importante para garantir a segurança dos turistas. Por exemplo no caso de um destino de neve, a previsão da sua existência é importante para a viabilidade dos negócios locais.

Com o aumento da população munida a agricultura afirma-se cada vez mais como uma área crítica. Estando muito vulnerável a eventos climáticos adversos e extremos, a agricultura é sem dúvida uma das áreas que mais beneficia com análise dos dados climáticos. Munidos das previsões de eventuais fenómenos destrutivos das colheitas como ocorrência de secas ou cheias, os produtores podem tomar medidas preventivas que minimizem o impacto negativo na produção agrícola. Com a boa exatidão dos resultados da *Big data* aplicada à previsão do tempo, os atores económicos podem estimar melhor os preços de compra e venda e os retalhistas conseguem reforçar os seus stocks de acordo com os acontecimentos que se espera que possam acontecer.

Quando a análise de dados meteorológicos através da *Big data* é aplicada à construção civil existem vidas que podem ser salvas e ferimentos que podem ser evitados nos funcionários das empresas desta área. Cada vez mais as empresas desta área estão consciencializadas de que o bem-estar dos seus funcionários é relevante para o sucesso empresarial e desta forma usam previsões para reduzir riscos. Um exemplo prático pode ser a utilização de gruas de elevada altitude em dias com ventos fortes. Se numa obra existe a necessidade de manipular os equipamentos anteriormente referidos e estão previstos ventos fortes, uma tempestade ou um tornado, então esse trabalho deverá ser evitado. A previsão de cheias em determinadas zonas geográficas pode ser um fator determinante para que seja dada uma atenção redobrada à construção de edifícios de modo que não ocorram deslizamentos de terras ou inundações.

Outra área que beneficia muito da *Big data* para prevenção de eventos climáticos extremos é o setor dos transportes. Com informações sólidas e precisas como a ocorrência de tempestades no mar ou de tornados, os tráfegos marítimos e aéreos são atempadamente alertados para situações perigosas adaptando-se de modo a não correrem riscos dado que o setor dos transportes é vital para a economia mundial. Com informações verossímeis as companhias aéreas podem reagendar voos que se realizariam em momentos para os quais está previsto mau tempo. Ao nível viário, tempestades de

neve e chuva em abundância podem ser previstas com dias de avanço de modo a alertar os automobilistas para terem cuidados acrescidos.

O trabalho refere alguns obstáculos inerentes à *Big data* e que condicionam a previsão de fenómenos que podem pôr em risco vidas humanas e as suas atividades. Por definição, a *Big data* lida com volumosas quantidades de dados pelo que podem surgir limitações no âmbito do armazenamento e processamento desses dados. No entanto outro dos problemas da área tem que ver com o facto de não existir um histórico de dados com qualidade e disponibilidade. As estações meteorológicas muitas vezes não possuem dados mais antigos que podem ser importantes para prever fenómenos climáticos extremos mais raros. Ao nível tecnológico é referido que os centros meteorológicos devem ser vistos como instalações críticas e tratadas como tal. Apostar na redundância de equipamentos informáticos e na fiabilidade do software devem ser encaradas como duas medidas chave para garantir disponibilidade dos sistemas e para que as previsões sejam confiáveis.

O trabalho conclui que é expectável que a recolha de dados se manifeste como uma variável de complexidade a ser trabalhada no futuro dado que muitas fontes de dados de meteorologia envolvem custos avultados. No entanto esses custos são inferiores ao que seriam causados por desastres naturais que a *Big data* e previsão meteorológica conseguem identificar, justificando o investimento nesta área. Em suma, a *Big data* aplicada à prevenção de desastres naturais mostra-se verdadeiramente útil permitindo antecipar problemas e minimizar os mesmos contribuindo para a segurança das pessoas e para o seu bem-estar.

O trabalho redigido por Cox et al (2018), tem como principal objetivo apresentar um modelo de previsão da trajetória dos furacões com origem no oceano Atlântico que todos os anos provocam avultados prejuízos materiais e ainda acarretam a perda de vidas humanas e ferimentos graves a muitas pessoas. Prevendo a trajetória dos furacões é possível implementar medidas preventivas como evacuação de aglomerados populacionais.

Começa-se por referir que a disponibilidade e a grande taxa de geração de dados, por exemplo por soluções de internet das coisas, é uma das justificações da utilização da *Big data* nos modelos de previsão da trajetória de furacões (MPTF). A utilização da *Big data* nesta área concentra-se em duas áreas distintas: previsão da intensidade do furacão e a previsão da trajetória do furacão (sobre a qual incide o estudo em causa).

Os dados sobre furacões e as suas trajetórias, que vão desde 1950 a 2000, são analisados com uma especial atenção num processo de *sequential mining* como é o caso do algoritmo *AprioriAll* que é o usado. Os autores apresentam o seu modelo a que deram o nome de *Weighted-Asset, Realistic-Distance Hurricane Trajectory Predictor* (WARD-HTP) e cujos dados usados provêm do HURDAT2, uma base de dados sobre furacões originários no oceano Atlântico e que está a cargo da *US National Oceanic & Atmospheric Administration*. Estes dados foram tratados sendo que os

autores apenas usaram dois campos com um relacionamento do tipo chave (Identificação do furacão) e valor (sequência da trajetória do furacão) por exemplo: AL081976 - ((26.0:-84.0), (25.3:-83.3), (24.7:-82.7))

O processo inicia-se com a seleção de registos da base de dados que vão sendo analisados e passando a níveis superiores de análise. Posteriormente com a plena utilização do algoritmo *AprioriAll* devem ter sido identificadas trajetórias padrão que satisfaçam alguns níveis estipulados de modo a garantir uma precisão do modelo a desenvolver. A fase seguinte consiste em gerar regras que possibilitam a identificação de associações de deslocamento dos furacões.

Relativamente aos dados das coordenadas de trajetória, é referido que se estes estiverem num estado de demasiada granularidade podem dificultar o todo o processo levando a um processo de “*discretization*” que consiste em transformar uma medida contínua numa medida discreta. Deste modo, a área geográfica em estudo (zonas do planeta adjacentes ao oceano atlântico), foram divididas em blocos de tamanho fixo de aproximadamente 500km por 500 km.

De modo a obter uma maior precisão do modelo proposto, é usada uma fórmula de distância real que dá pelo nome de *harvestine distance*, cujo objetivo é calcular a distância entre dois pontos numa esfera, sendo que pode ser aplicada ao nosso planeta.

Foi associado um peso aos dados para que os dados mais antigos tenham menos influência que os mais recentes. Desta forma, padrões obtidos por dados mais recentes não precisarão de se repetirem muitas vezes para serem identificados ao contrário do que acontece com dados mais antigos.

Depois de intensamente testado com dados de 2011 a 2015, o modelo implementado em Python, obteve uma precisão máxima de 75%, sendo superior a outros modelos do género, e como trabalho futuro, os autores identificaram a incorporação de outros atributos/campos de input como a distribuição da pressão atmosférica ou a temperatura da água.

Segundo o artigo escrito por Madan et al (2018), a população indiana diariamente sofre com um mau clima inconstante, poluição e as suas consequências. Agricultores em todo o mundo estão a enfrentar dificuldades derivadas de mau tempo. Daí, a previsão do clima ser cada vez mais importante para tentar de alguma forma ajudar a reduzir algumas destas consequências relatadas anteriormente.

Big data contém uma gigantesca quantidade de informação podendo esta ser estruturada, semiestruturada ou não estruturada, sendo uma das razões pela qual é extremamente difícil processar e guardar toda esta informação. Mas estes dados podem ajudar a prevenir o clima de uma determinada região. Por isso, neste artigo é então tratada uma das formas de prever o clima através da análise de grandes quantidades de dados (*Big Data*), usando *machine learning* e *data mining*, juntamente com técnicas de regressão linear, e *vector machine*. Estas técnicas permitem a extração de apenas dados mais importantes de uma *collection* de *Big Data*, mesmo estas bases de dados sendo

tão difíceis de tratar. É possível prever o clima com base na concentração de determinados componentes atmosféricos, como o dióxido de carbono, nitrogénio, entre outros. Este artigo concentra-se na previsão nomeadamente na área de *Delhi*. As principais técnicas de *data mining* utilizando *machine learning* são Classificação, Agrupamento, Associação e Regressão. A questão da medição do clima incorpora a medição da temperatura, chuva, névoa, ventos, e tempestade e assim por diante. Os sensores climáticos reúnem-se informação de forma consistente em numerosas áreas e reúnem uma quantidade de informação tremenda. A antecipação climática é um teste importante, uma vez que é difícil prever a qualidade do ar num futuro próximo, uma vez que o conjunto de dados atmosféricos são inconstantes e mudam de dia para dia.

A informação utilizada neste estudo é da autoria do IMD, ou seja, *India Meteorological Department*, com uma *collection* de *Big Data* de dados semi-estruturados. Para uma melhor compreensão do que é realmente a *Big Data*, os autores fazem uma breve menção de como esta surgiu, do que esta representa, assim como algumas das suas especificações como por exemplo, as 8 qualidades da *Big Data*, ou seja, os 8V's.

Para a realização deste artigo foram analisados diversos outros artigos. Grande parte destes artigos estão relacionados com o uso de *Big Data* e a análise de dados climáticos. Para este efeito foram descritos diversos sistemas que dão uso a tecnologias de inteligência artificial. As redes neuronais, uma das principais componentes da inteligência artificial foi uma das tecnologias mais evidenciadas nos artigos apresentados, tendo estas demonstrado um desempenho superior relativamente a modelos de regressão linear e funcional, dado que conseguem capturar condições indiretas de padrões climáticos anteriores, com base em dados de histórico.

Os autores também fazem menção ao *Hadoop*. Esta é uma ferramenta que permite a filtragem de spam, análise de dados, entre outros e, é usado por empresas como Facebook e a Yahoo e pode ser de grande utilidade à *Big Data*.

Posteriormente é apresentada um breve *survey* onde os autores explicitam vários outros artigos relacionados com o tema em questão. Alguns dos artigos apresentados falam de temas como por exemplo um microcontrolador e nas suas aplicações para *machine learning*, ou cálculos para sistemas estatísticos avançados, ou ainda a possibilidade de usar várias fontes de informação para melhorar as capacidades de proteção de infraestruturas críticas, com ênfase em técnicas como a de análise de dados.

No artigo, seguidamente é apresentada o trabalho a que os autores se propuseram, que é nomeadamente analisar os dados climáticos extremos de *Delhi* referidos em cima, que vão desde 1996 a 2017. Estes dados encontram-se escritos em 9 estados que foram divididos em 4 categorias como razoavelmente nublado, muito nublado, tempestade e precipitação. Posteriormente foi apresentado um método de regressão linear para prever a temperatura máxima e mínima, assim como

também um método *vector machine* para limitar a função de erro sendo que para este modelo são utilizados dados mais antigos para treinar o modelo de previsão e, dados mais recentes para testar a sua precisão.

Para os autores obterem os resultados desejados começaram primeiro realizaram *Map Reduce* com uso de *Big Data (Hadoop)*, depois descrevem a forma como realizaram a regressão linear e suporte do método *vector machine*.

Posto isto, os autores chegaram a algumas conclusões, nomeadamente a que embora os métodos de previsão climática tradicionais ainda superem o método de previsão climática apresentado por este artigo numa fase inicial, nos dias seguintes o erro demonstrado pelo método apresentado pelos autores diminuiu bastante o que sugere que este método pode então superar os métodos tradicionais se, existir por exemplo recurso a mais dados.

Os autores Khan et al (2020) começam por evidenciar que, devido a mudanças climáticas e outros fatores ambientais, que o número de cheias tem vindo a aumentar ao longo dos anos. Estes aumentos têm vindo a ter um grande impacto nas vidas humanas. Através do uso combinado de *machine learning*, dados de sensores IoT é possível prevenir os níveis de inundação dos rios de forma precoce, de forma que por exemplo as equipas de resgate possam ter uma resposta mais eficaz.

A previsão de uma cheia é difícil de estimar, devido às variações de variáveis temporais e espaciais e, a atividade do fluxo do rio é influenciada por diversos aspetos, tais como, topografia do leito do rio, chuvas e outras características climáticas.

Os autores referem também a existências de duas abordagens diferentes à previsão de uma cheia. A primeira, através do uso de modelos matemáticos, consegue compreender a hidrodinâmica do fluxo da água. Neste caso é necessário um grande fluxo de dados de entrada e, a seleção dos parâmetros necessários precisam de ser selecionados com cuidado, ou seja, quando os dados são restritos, estes tipos de modelos não funcionam bem. Já a segunda abordagem recorre a dados para a previsão do fluxo do rio através do recurso a redes neuronais e algoritmos de *machine learning*, que permitem interpretar as propriedades lineares e não lineares dos dados e, eliminar a necessidade de fornecer quaisquer suposições e assim prever o fluxo do rio.

Os autores analisaram diversos estudos relativos à previsão de cheias com recurso a algoritmos de *machine learning*, onde alguns falam de regressão linear, outros utilizaram nas suas pesquisas algoritmos de *machine learning*, e outros utilizaram até dados hidráulicos e hidrológicos para produzir primeiramente mapas de risco e só então utilizaram *machine learning*.

Com a realização deste artigo os autores pretendem: propor uma abordagem baseada em *Big Data* para detetar a gravidade do nível da cheia a ocorrer com recurso a sensores IoT; utilizar PCA (análise dos principais componentes), para ajudar a identificar as características mais importantes e

respetivos pesos nos componentes principais; e investigar a representação das classes desequilibradas nos dados e dar uso à técnica *Sintetic Minority Oversampling* para compreender quais os impactos da performance da técnica anterior e algoritmos *Random Forest*.

Posteriormente os autores começam por apresentar a metodologia proposta baseada no processamento de dados usando *machine learning*. Os dados são do *Environment Agency Dartmouth Flood Observation*, com uma base de dados de pelo menos 31 anos, que consiste em dados recolhidos por vários sensores de todo o mundo relativamente a cheias e outros indicadores importantes para estas. No estudo são referidas 3 classes correspondendo a primeira a níveis normais, a segunda a anormais e a terceira a níveis de alto risco de água do rio.

Sendo a quantidade de dados muito extensa estes foram pré-processados. Posteriormente, os autores deram uso à técnica *Oversampling*, de modo que as diferentes classes passassem a ter uma representação mais justa e não tão discrepante (C1 com 892, C2 com 918 e, C3 com 854 amostras).

Neste artigo foram utilizados dois algoritmos de *machine learning* para analisar os dados obtidos anteriormente, nomeadamente o algoritmo *Support Vector Machine* (SVM), usado para resolver por exemplo problemas complexos de regressão linear e o *Random Forest*, que permite gerar previsões confiáveis neste caso, da gravidade da inundação. Na aplicação destes algoritmos foram realizadas diversas avaliações relativamente ao desempenho destes dois algoritmos. Consoante os resultados destes testes foram seleccionadas as configurações dos parâmetros para resultados mais fidedignos.

Quanto aos resultados, foi possível verificar que tanto um algoritmo como o outro tiveram desempenhos semelhantes no que toca a precisão da representação dos dados com cerca de 92%, aquando da utilização dos 11 atributos (a sua totalidade). Posteriormente foram realizados vários testes mediante a precisão dos algoritmos na qual se pode tirar algumas conclusões, tais como, quando o algoritmo SVM utiliza menos atributos, nomeadamente 7 dos mais importantes, a sua eficácia é mais reduzida. Contrariamente, o *Random Forest* ao utilizar essa quantidade mais reduzida de atributos, verificou-se uma eficácia ainda maior que o seu primeiro valor com a totalidade dos atributos, ou seja, a quantidade de atributos utilizada pelos algoritmos é de uma extrema importância para o seu sucesso final. A seleção dos atributos mais importantes influencia o resultado da eficácia dos algoritmos principalmente no algoritmo *Random Forest*.

Depois de todos estes testes, os autores concluíram primeiramente que, a representação das amostras equilibradas origina uma percentagem de eficácia nos dois algoritmos maior. O algoritmo de *machine learning*, *Random Forest* alcançou uma eficácia de 93% quando os testes foram realizados com 10 atributos ao invés da totalidade deles (11). Posto isto pode-se verificar que todas as premissas referidas anteriormente aliadas aos algoritmos de *machine learning* podem ser uma ajuda a realizar uma previsão bastante precisa dos níveis de risco de cheia num rio, neste caso.

O trabalho de Li et al (2020), foca-se num sistema que permita alertar antecipadamente eventos climáticos extremos que possam pôr em causa a integridade e normal funcionamento da rede de distribuição elétrica.

Com as alterações climáticas, os fenómenos climáticos extremos têm vindo a aumentar e com eles a preocupação das entidades responsáveis pela distribuição de energia elétrica faz a esses fenómenos. A rede elétrica e os seus equipamentos encontram-se expostos a todos os fenómenos atmosféricos extremos e não extremos dado que se localizam muitas vezes em campo aberto.

Vento, trovoadas, geadas, chuvas fortes que provocam deslizamentos de terras são verdadeiros desafios, segundo os autores, para a rede elétrica e os respetivos equipamentos. Combinando os dados recolhidos na rede elétrica com dados climáticos é possível desenvolver um sistema que antecipe situações climáticas extremas permitindo ainda que responsáveis da rede elétrica possam construir melhores planos de prevenção e resposta a essas situações. Desta forma os efeitos prejudiciais podem ser minimizados evitando “apagões” em grandes aglomerados populacionais. Os autores justificam o seu estudo referindo o facto de existir uma lacuna significativa entre os dados meteorológicos usados provenientes de departamentos meteorológicos que possuem as estações de recolha de dados a grandes distâncias das linhas e equipamentos de distribuição elétrica.

Os autores apresentam um processo de alerta antecipado de eventos meteorológicos onde ocorre a recolha em tempo real de vários tipos de dados sendo depois analisados permitindo criar previsões meteorológicas que dão informações sobre quando e como ocorrerão situações extremas permitindo emitir comandos de emergência de prevenção. Entendendo quais os prejuízos que aconteceriam caso nada fosse feito é possível melhorar o plano de prevenção de desastres que visam minimizar o impacto de fenómenos naturais na rede elétrica. Todo este processo está assente numa camada a que os autores dão o nome de *Foundation Support Platform*, camada essa que é composta por recursos computacionais desde servidores de processamento e/ou armazenamento e ainda instalações de redes de comunicação. Em caso de existir um alerta de evento climático extremo, essa informação é comunicada às autoridades governamentais responsáveis pela proteção civil.

O trabalho realça o papel do GIS (*Geographic Information System*), que consiste numa tecnologia de gestão de informação espacial, para lidar com quantidades massivas de dados para previsão meteorológica e de desastres.

O sistema a implementar, afirmam os autores, lidará com dados não estruturados, estruturados, recolhidos em tempo real ou não como por exemplo dados de radar, dados históricos, registo de nuvens e furacões. Este sistema encontra-se segmentado em 5 camadas. A *Presentation Layer* que apresenta os dados tratados por meio de gráficos, por exemplo. A *Application Layer* encarrega-se da monitorização de alertas gerados e dos equipamentos da rede elétrica. A camada *Data Processing Layer* é a responsável pelo processamento dos dados e pelo processo de *Big data* enquanto a *Data*

Storage Layer engloba o armazenamento dos dados provenientes de várias fontes. Por fim a *Software and Hardware Infrastructure Layer* engloba os recursos necessários a todas as outras camadas.

De modo a lidar com o grande volume de dados, o trabalho refere a utilização de uma arquitetura distribuída para armazenamento e processamento que também assegura a disponibilidade de todo o sistema. Em termos de infraestrutura física, o core do sistema (servidores), encontrar-se-ia na sede da produtora elétrica e a comunicação dos dados seria feita usando os protocolos HTTP e TCP.

Por fim, os autores referem que o sistema proposto servirá para identificar e mitigar o efeito de fenómenos climáticos como é o caso de deslizamentos de terra, chuvas repentinas, nevoeiro, tufões e temperaturas muito baixas ou muito altas.

4. DISCUSSÃO

O primeiro artigo por nós analisado defende a utilização da *Big data* como uma das técnicas de análise e previsão meteorológica que pode ajudar evitar e prevenir algumas das consequências de determinados fenómenos climáticos podem ter na vida humana. Ele faz especificamente menção a algumas áreas que podem beneficiar com esta técnica nomeadamente a segurança da população, turismo, agricultura, construção civil, transportes, entre outros. Neste artigo também são referidas algumas das limitações que vêm com a utilização da *Big data*, nomeadamente a grande quantidade de dados ou até mesmo a falta deles, o que torna a previsão menos precisa (Jain & Jain , 2017).

Os seguintes quatro artigos já todos referem métodos ou sistemas de previsão, grande parte deles utiliza dados recolhidos por sensores de IoT, assim como, muitos deles referem a maneira em como trabalharam *Big Data* de modo a aplicarem um determinado algoritmo (grande parte refere algoritmos de *machine learning*) e assim preverem o clima ou algum evento mais específico, nomeadamente:

Em Li et al (2020), os autores referem um sistema de previsão climática de eventos que possam danificar sistemas de rede e distribuição elétrica, de forma a que as entidades responsáveis possam atuar de modo a prevenir algumas falhas daí decorridas. Este sistema utiliza a *Big data*. Os dados são recolhidos por sensores e tratados com recurso a uma tecnologia, nomeadamente o GIS que consegue tratar de quantidades massivas de dados. O sistema referido pelos autores de previsão encontra-se estruturado em camadas, dando uso aos protocolos de rede HTTP e TCP para fazer as comunicações necessárias.

O artigo terceiro apresenta também um modelo de previsão da trajetória dos furacões com recurso a sensores IoT e Big data sendo os referidos dados analisados com recurso ao algoritmo *AprioriAll*. O modelo também recorre à linguagem *Python* e tem uma precisão de cerca de 75% (Cox et al, 2018).

Já Madan et al (2018), apresenta um método de previsão do clima, mas neste caso utiliza *Big data* e *machine learning*, *data mining*, técnicas de regressão linear, entre outras. Este sistema, no final dos seus testes apresentou uma percentagem de erro menor, o que sugere que futuramente este método de previsão, se houver recurso a uma maior quantidade de dados, pode superar os métodos tradicionais de previsão do clima.

Os autores Khan et al (2020), defendem um método de previsão de cheias e consequente risco, usando *machine learning*, com recurso a dados recolhidos por sensores IoT, combinando redes neuronais e algoritmos de *machine learning*. Neste artigo foram evidenciados cerca de 2 algoritmos. A cada um dos algoritmos de previsão foram realizados vários testes sendo que, o teste que deu uma percentagem maior conjunta teve uma percentagem de eficácia de cerca de 93%, o que é bastante bom.

5. CONCLUSÃO

Com a realização deste trabalho conseguimos concluir que muitos dos artigos por nós analisados referem que a incerteza do clima atualmente é afetada por determinados fatores e, que a vulnerabilidade não só dos humanos como de outras seres vivos ou como por exemplo sistemas de rede e distribuição elétrica são algumas das razões pela qual é de uma extrema importância a previsão do clima, mais concretamente a previsão climática de eventos extremos.

Grande parte destes artigos referiram o quanto a *Big data* pode ajudar para este fim, nomeadamente em conjunto com sensores de IoT a recolher mais dados para estas bases de dados e para serem analisados, alguns dos artigos referiram algoritmos ou métodos de seleção dos dados, e posteriormente, também grande parte dos artigos analisados referiram o trabalho em conjunto com um ou mais algoritmos de *machine learning* para a previsão meteorológica.

Foi-nos possível verificar inclusive, que se estas bases de dados não fossem suficientemente completas e munidas de amostras suficientes, por muito bons e eficazes que os algoritmos fossem de análise e previsão do clima, estes iriam sempre ter um resultado com uma menor eficácia do que se isto não acontecesse.

Os resultados da eficácia de grande parte dos métodos de previsão apresentados pelos artigos analisados obtiveram uma percentagem de eficácia bastante positiva, ultrapassando até num caso (e depois de alguns testes) um sistema de previsão do clima profissional.

Concluindo, a *Big data* tem cada vez mais um papel de referência nas nossas vidas, permitindo neste caso da previsão climática e, mais em concreto, da previsão de eventos climáticos extremos, o que pode ajudar a prevenir e/ou evitar grades catástrofes.

REFERÊNCIAS

- Cox, T. S., Hoi, C. S. H., Leung, C. K., & Marofke, C. R. (2018). An accurate model for hurricane trajectory prediction. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 02, 534–539.
- Jain, H., & Jain, R. (2017). Big data in weather forecasting: Applications and challenges. 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), 138–142.
- Khan, W., Hussain, A. J., Alaskar, H., Baker, T., Ghali, F., A-Jumeily, D., & Al-Shamma'a, A. (2020). Prediction of flood severity level via processing IoT sensor data using a data science approach. *IEEE Internet of Things Magazine*, 3(4), 10–15. <https://doi.org/10.1109/iotm.0001.1900110>
- Li, D., Gong, Y., Shen, S., & Feng, G. (2020). Research and design of meteorological disaster early warning system for power grid based on Big data technology. 2020 Asia Energy and Electrical Engineering Symposium (AEEES), 658–662.
- Madan, S., Kumar, P., Rawat, S., & Choudhury, T. (2018). Analysis of weather prediction using machine learning & Big data. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), 259–264.