1. Introdução

Apesar dos altos investimentos dos países, milhões de hectares florestais são destruídos pelo fogo em todo o mundo a cada ano. A maior parte deste incêndios são causados pelas pessoas, intencionalmente ou não, e por relâmpagos.

A combinação de um clima quente e seco, com a primeira fagulha e adicionando o combustível para alimentar as chamas (madeira inflamável, arbustos e folhagens secas), e vento forte, pode levar a graves desastres.

De 1980 a 2005, mais de 2,7 milhões hectares (ha) de área florestal foram destruídas em Portugal. Os incêndios de 2003 e 2005 foram devastadores, afetando 4,6% e 3,1% do território e deixando 21 e 18 pessoas mortas respectivamente.

As ferramentas mais utilizadas para detectar o incêndio rapidamente são:

* baseado em satélite
* scanners de infravermelho / fumaça
* sensores locais (por exemplo, meteorológicos).

Os satélite têm custos de aquisição, atrasos de localização e a resolução não é adequada para todos os casos e os scanners têm um alto custo de equipamento e manutenção. Já os sensores locais coletam dados climáticos em tempo real, como temperatura e umidade do ar com baixos custos.

Várias técnicas de mineração de dados, são aplicadas para a detecção de incêndio. Por exemplo, redes neurais (NN) para prever a ocorrência de incêndios causados por humanos. Scanners infravermelhos e NN foram combinados para reduzir alarmes falsos de incêndio florestal com um sucesso de 90%. Um agrupamento espacial para detectar pontos de incêndio florestal em imagens de satélite.

O Índice Climático de Incêndio Florestal (FWI) ou O cálculo dos componentes é baseado em observações diárias consecutivas de temperatura, umidade relativa do ar, velocidade do vento e chuvas de 24 horas. Os seis componentes padrão fornecem classificações numéricas de potencial relativo para incêndios florestais.

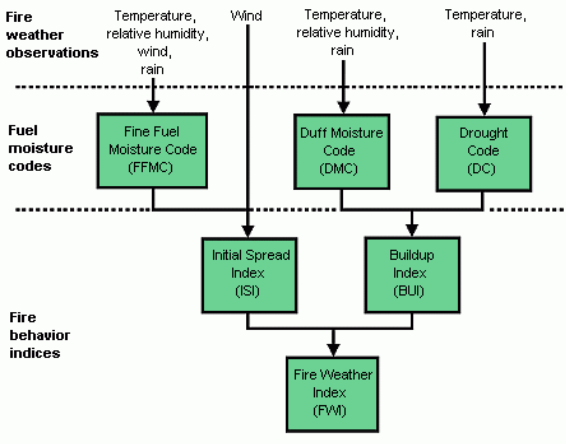
O Índice Climático de Incêndio Florestal (FWI) é o sistema canadense de classificação de risco de incêndio e inclui seis componentes (Figura 1).

Os três primeiros estão relacionados às condições climáticas propícias para gerar o combustível:

* O **Código Umidade Combustível FFMC** influencia a **propagação de ignição** e incêndio.
* O **Código de umidade (DMC)**  e **Código da Seca (DC)** afetam **intensidade do fogo**.

Os últimos representam os índices de comportamento de fogo:

* O **Índice Propagação Inicial (ISI),** é uma pontuação que **se correlaciona com a propagação da velocidade do fogo.**
* O **Índice de Construção (BUI)**  representa a **quantidade de combustível** disponível.



O Índice Climático de Incêndio Florestal (FWI) é um indicador da intensidade do fogo e combina o Índice Propagação Inicial (ISI) e o Índice de Construção (BUI). Apesar de diferentes escalas sejam usadas para cada um dos elementos da FWI, seus altos valores sugerem condições de queima mais intensas.

2. Base de dados

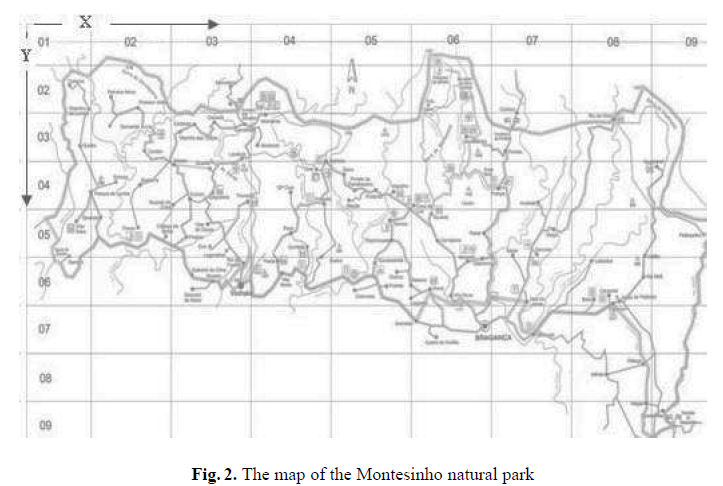
Os dados utilizados nos experimentos foram coletados no Parque Natural de Montesinho, cerca de 1000 metros de altitude e temperatura média anual de 8 a 12◦C, na região nordeste de Portugal e possui uma alta flora e diversidade de fauna (Figura 2).

Os dados utilizados nos experimentos foram coletados de janeiro de 2000 a dezembro de 2003 e foram construídos com duas fontes. A primeira base de dados foi recolhida pelo inspetor responsável pelas ocorrências de incêndio de Montesinho.

Eram registradas a hora, a data, a localização espacial em uma grade de 9 × 9 (eixos x e y do Figura 2), o tipo de vegetação envolvida, os seis componentes do sistema FWI e a área total queimada.

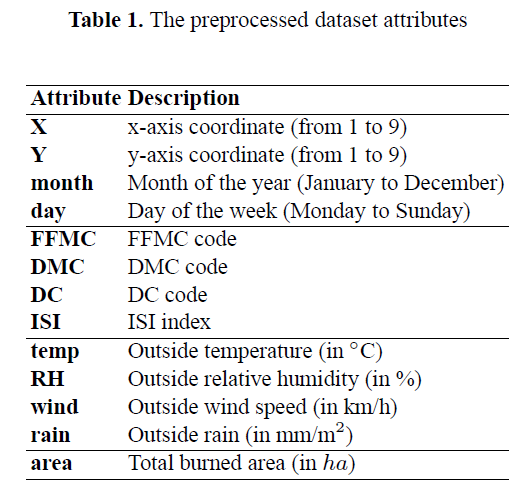
O segundo banco de dados foi coletado pelo Instituto Politécnico de Bragança, contendo várias observações meteorológicas (por exemplo, velocidade do vento) que foram registradas com um período de 30 minutos por uma estação meteorológica localizada no centro do parque de Montesinho.

Os dois bancos de dados foram unidos em um único conjunto de dados com um total de 517 entradas.



A tabela 1, apresenta duas características geográficas, os valores dos eixos X e Y em que ocorreram os incêndios e as variáveis ​​temporais mês e dia da semana nas primeiras quatro linhas. As condições meteorológicas médias mensais são bastante distintas, enquanto o dia da semana também pode influenciar os incêndios florestais (por exemplo, dias de trabalho vs fins de semana), uma vez que a maioria dos incêndios tem uma causa humana.

Em seguida, os quatro componentes do FWI afetados diretamente pelas condições climáticas, são quatro as variáveis ​​meteorológicas ( chuva, vento, temperatura e umidade). Existem 247 amostras com valor zero para o atributo “area”, isto significa que uma área menor que 1ha / 100 = 100m2 foi queimada.

****

Neste caso os valores são **registros instantâneos**, conforme dados pelos sensores da estação quando o incêndio foi detectado. Com exceção da variável “chuva”, que é medida a precipitação (qualquer tipo de fenômeno relacionado à queda de água do céu, neve, chuva e chuva de granizo)acumulada nos últimos 30 minutos.

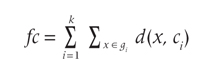
3. Algoritmos para o problema

Originalmente, esse problema foi modelado como uma tarefa de regressão, com o objetivo de prever pequenos incêndios, que constituem a maioria das ocorrências. A desvantagem é a menor precisão de previsão para grandes incêndios. E a meta neste trabalho foi realizar uma análise a fim de encontrar correlações utilizando o gráfico de calor e posteriormente a medir a similaridade entre os registros no conjunto de dados meteorológicos utilizando o algoritmo de agrupamento K-médias.

O parâmetro k corresponde ao número de grupos desejados, dividindo o conjunto de n objetos em k grupos, de forma que seja possível notar diferença entre os grupos e semelhança entre os objetos de um mesmo grupo. O centróide pode ser um valor médio dos objetos de um grupo ou aleatoriamente, selecionando-se alguns dos objetos da própria base de dados.

Calculando a distância de cada objeto para cada um dos centróides, teremos o valor mais próximo de cada objeto para cada um dos centróides e a este, serão atribuídos àqueles objetos. Através da média dos objetos pertencentes a cada centróide, um reposicionamento dos centróides é calculado e um novo grupo de objetos é gerado.

Quando não existe alterações nos centróides, algoritmo estabiliza e assim não é possível trocar os objetos de grupos. Tal processo equivale à soma dos erros quadráticos intragrupos; ou seja, a soma da distância de cada objeto ao centróide do grupo ao qual pertence:



onde fc é a função de custo (tempo de processamento) da base, x é um objeto qualquer da base, ci é o centróide do grupo gi, e d( x, ci) é a distância entre o objeto e o centróide do grupo.

O primeiro critério de parada é baseado na função de custo, que interrompe o algoritmo quando não há mudança no valor da função entre iterações, indicando que não houve mudança no posicionamento dos centróides e o segundo é o número máximo de iterações que garante a parada do algoritmo caso a função de custo ainda apresente pequenas variações.

Fonte da base: <http://archive.ics.uci.edu/ml/datasets/Forest+Fires>

[Cortez e Morais, 2007] P. Cortez e A. Morais. Uma abordagem de mineração de dados para prever incêndios florestais usando dados meteorológicos. Em J. Neves, MF Santos e J. Machado Eds., Novas Tendências em Inteligência Artificial, Anais da 13ª EPIA 2007 - Conferência Portuguesa de Inteligência Artificial, Dezembro, Guimarães, Portugal, pp. 512-523, 2007. APPIA, ISBN -13 978-989-95618-0-9. Disponível em: http://www3.dsi.uminho.pt/pcortez/fires.pdf

<http://www.aldeiasportugal.pt/sobre/78/#.W_qT2uhKjIU>

<https://super.abril.com.br/mundo-estranho/como-surge-um-incendio-florestal-e-como-combate-lo/>

DANIEL GOMES FERRARI,LEANDRO NUNES DE CASTRO SILVA. Introdução a mineração de dados (Locais do Kindle 3051-3057). Editora Saraiva. Edição do Kindle.