

Instituto de Educação Superior de Brasília - IESB (2022)  
Unidade de Pesquisa: CD e IA  
Equipes: Graduação e Pós-graduação  
Responsável: Prof. Sérgio Côrtes



CIÊNCIA DE DADOS E INT. ARTIFICIAL - IA

---

# **Análise de fenômenos naturais e implementação de um observatório sismológico**

---

**Victor Augusto Souza Resende**

1 de março de 2022

## **Resumo**

Terremotos são tremores que ocorrem na crosta terrestre. Esse texto visa a análise dos dados referentes aos sismos, dos quais foram coletados por meio do Serviço Geológico dos Estados Unidos (USGS). Para tal, são utilizadas diversas técnicas de estatística exploratória, inferência e modelagem de dados. Dessa forma, implementou-se um portal interativo e dinâmico, denominado Observatório Sismológico, do qual tem como objetivo a disponibilização visual dos dados referentes aos tremores nos filtros aplicados. Por fim, o portal traz a possibilidade da predição da magnitude de um tremor dado um valor de longitude e profundidade do epicentro determinado pelo usuário.

**Palavras-chave:** Terremotos, fenômenos naturais, sismologia, análise de dados, mineração de dados

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	Motivação . . . . .	4
1.2	Objetivos . . . . .	4
1.3	Trabalhos relacionados . . . . .	4
1.4	Composição do texto . . . . .	5
<b>2</b>	<b>Fundamentação Teórica</b>	<b>6</b>
2.1	Conceitos sismológicos . . . . .	6
2.2	Intensidade . . . . .	6
2.3	Placas tectônicas . . . . .	7
2.4	Escalas . . . . .	8
2.5	Medição e monitoramento . . . . .	9
2.6	Conceitos estatísticos . . . . .	10
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>11</b>
<b>4</b>	<b>Metodologia</b>	<b>12</b>
4.1	Dicionário de dados . . . . .	12
4.2	Arquitetura da solução . . . . .	12
4.3	Ferramentas utilizadas . . . . .	13
<b>5</b>	<b>Desenvolvimento</b>	<b>14</b>
5.1	Extração e manipulação dos dados . . . . .	14
5.2	Análise exploratória . . . . .	14
5.3	Predição da magnitude — Floresta aleatória . . . . .	19
5.4	Observatório sismológico . . . . .	22
5.5	Pontos de melhoria . . . . .	24
<b>6</b>	<b>Conclusão</b>	<b>25</b>
	<b>Références</b>	<b>26</b>

# 1 Introdução

A **Ciência de Dados** é uma área nova de ciência, multidisciplinar, que se baseia na integração das áreas de matemática, estatística e computação. Em termos simples, resume-se na exploração e análise de dados visando à extração de informações e conhecimento a partir dos dados. A Ciência dos Dados em muitos aspectos é uma consequência da necessidade de analisar grandes bancos de dados (Big Data), até pouco tempo conhecidos como Very Large Database (VLDB), que possuem características de grandes volumes de dados, alta velocidade de geração e armazenamento de novos dados, variedade de tipos de dados, alta qualidade nos dados e o processamento de seus dados agregam valor nas organizações. A Ciência de Dados é um campo interdisciplinar que exige a formação de um novo profissional que possua as habilidades e competências no uso da matemática, da estatística e da computação, aplicados na extração de informações e conhecimentos dos dados ([Han et al. \[2012\]](#)).

A sismologia visa o estudo dos sismos (ou terremotos) e, genericamente, dos diversos movimentos que ocorrem na superfície do globo terrestre. Esta ciência busca conhecer e determinar em que circunstâncias ocorrem os sismos naturais assim como suas causas, de modo a prevê-los em tempo e espaço. Portanto, por meio dessa ciência, é possível analisar dados gerados de diversos observatórios sismológicos e sensores sismógrafos a fim de entender os tremores terrestres, as causas e impactos diretos na sociedade, havendo até a possibilidade de prevê-los em alguns casos dependendo dos dados gerados.

O objetivo do observatório sismológico é a disponibilização de uma visualização dos tremores ocorridos conforme os filtros realizados pelo usuário. Da mesma forma, o ambiente também apresenta uma etapa onde é possível fazer a predição dos sismos por meio da longitude e profundidade do epicentro. Esse estudo considera os principais textos publicados sobre o tema de predição de terremotos, por [Mondol \[2021\]](#) e [Geller et al. \[1997\]](#).

Para a confecção do observatório sismológico, fez-se a utilização de diversas ferramentas de mineração, coleta, criação de modelos preditivos e implementação da visualização e predição dos dados. Os dados utilizados para as análises a seguir são oriundos de uma API disponibilizada pelo Serviço Geológico dos Estados Unidos (USGS) <sup>1</sup>. Assim, o estágio de coleta dos dados foi realizado por meio da linguagem de programação Python, da qual se arquitetou uma lógica para a extração das variáveis pertinentes e seus respectivos dados. Por fim, a etapa de criação de modelos de predição e implementação do observatório sismológico se desenvolveram por meio das bibliotecas Python citadas ao decorrer do projeto.

Portanto, todas as análises consideram os textos citados dos quais foram essenciais para a implementação do portal sobre o tema. Vale ressaltar que todo código-fonte são de autoridade do autor desse artigo, entretanto é disponibilizado ao público em repositório GitHub. Para acessar o portal do Observatório Sismológico, basta acessar a seguinte URL: .

---

<sup>1</sup>USGS: Serviço Geológico dos Estados Unidos

## 1.1 Motivação

Com o passar dos anos e o avanço da tecnologia, houve a criação e implementação de sensores em locais com risco de desastres naturais para a verificação de riscos e coleta dos dados. Dessa forma, uma vasta quantidade de dados é gerada diariamente, principalmente quando há situações de tremores, seja em razão de terremotos, erupções, ou até mesmo de ações humanas como acontece em alguns tipos de explosões. Portanto, a criação de um observatório sobre tremores e a previsão da magnitude de determinada vibração terrestre, dado a longitude e profundidade do tremor, torna-se interessante para o monitoramento de tais fenômenos.

## 1.2 Objetivos

Este projeto visa a criação de uma solução arquitetada de ponta a ponta, desde a coleta dos dados até a implementação da visualização e modelo de previsão interativo com o usuário. Por meio do histórico dos dados gerados até então, será possível visualizar as regiões onde tremores são mais comuns, as magnitudes mais altas registradas até então, e diversos outros fatores como, por exemplo, encontrar os tipos de terremotos que mais ocorreram na data especificada.

A implementação de um modelo de previsão, por meio da floresta aleatória, será interativa e poderá demonstrar ao usuário quais as regiões do globo terrestre estão mais propensas a ocorrer tremores, bem como a magnitude que o fenômeno pode alcançar em tal área. Portanto, o observatório deverá consumir os dados apresentados na API do Serviço Geológico dos Estados Unidos, gerando uma página de visualização dos dados em um gráfico de mapa, uma página com o modelo de previsão responsivo ao usuário, resultando em um tipo de monitoramento completo dos fenômenos causadores de terremotos.

## 1.3 Trabalhos relacionados

Esta seção visa apresentar textos relacionados com os tópicos abordados na criação do texto.

[Motoki et al. \[2013\]](#) escreve sobre a implementação do Observatório Sismológico (SIS) da Universidade de Brasília (UnB), do qual faz a captação, por meio de suas estações sismográficas, de sinais de tremores de terra. Os dados registrados são encaminhados, via satélite, das estações para a central do Observatório, no campus Darcy Ribeiro. No local, um software detecta automaticamente a ocorrência, acompanhada por pesquisadores e técnicos, que processam e analisam as informações recebidas para, em seguida, divulgá-las à população por meio do portal portalsis. O observatório possui um convênio para apoio à RSBR<sup>2</sup>.

[Mondol \[2021\]](#) reflete sobre um método viável e preciso para a previsão de terremotos tem o potencial de salvar incontáveis vidas humanas. Dessa forma, o autor aborda diversos métodos estatísticos para prever a magnitude e profundidade dos terremotos. Os desempenhos são comparados para encontrar o modelo mais eficaz. Por fim, chega-se a conclusão que é muito difícil prever com precisão a magnitude de terremotos, no entanto, o texto cita o bom desempenho dos modelos de regressão polinomial mostra os melhores resultados globais e as florestas aleatórias, das quais foram incrivelmente eficazes na previsão da profundidade de um terremoto.

[Geller et al. \[1997\]](#) aborda no texto se a hora, localização e a magnitude dos terremotos futuros podem ser previstas de forma confiável e precisa, em sua perspectiva, a resposta de Geller et al. é “não”. Citando resultados recentes da física de sistemas não lineares “teoria do caos”, argumentam que qualquer pequeno terremoto tem alguma hipótese de entrar em cascata num grande evento. Conforme a investigação citada pelos autores, se isto acontece ou não depende de detalhes incomensuráveis das condições no interior da Terra. Os terremotos são, portanto, intrinsecamente imprevisíveis.

---

<sup>2</sup>RSBR: Rede Sismográfica Brasileira

## 1.4 Composição do texto

Na introdução, seção 1, estão evidenciados todos os escopos da confecção deste projeto. Portanto, é onde foram discutidos tópicos como as definições iniciais, a motivação e o objetivo da criação do observatório de visualização e predição de fenômenos naturais causadores de terremotos.

Na seção 2 encontra-se a fundamentação teórica, da qual tem como objetivo trazer outros pontos de vista sobre o assunto estudado, enriquecendo o texto. Da mesma forma, nessa seção encontram-se as considerações a serem feitas para os conceitos utilizados no desenvolvimento do texto.

A seção 3 conta com a revisão bibliográfica. Essa seção demonstra como os textos e trabalhos relacionados foram essenciais nas tomadas de decisão, análise dos dados e implementação do portal.

Na etapa da metodologia, encontrada na seção 4, é onde apresentam-se possíveis respostas às hipóteses e a implementação do observatório sismológico, detalhando assim todo o processo de maneira aprofundada em relação ao processamento, tratamento, análise e predição dos dados coletados.

O desenvolvimento do texto na seção 5, demonstra as análises feitas, das quais englobam a análise exploratória dos dados do período de 15/12/2021 a 10/06/2022 com terremotos de no mínimo 2 graus na escala Richter. Essa etapa do texto aborda sobre a dificuldade na predição da magnitude dos terremotos por meio do modelo de Random Forest, ou em português, Floresta Aleatória. Por fim, esse tópico cita alguns pontos de melhoria mapeados para futuras implementações.

Por fim, a conclusão apresentada na seção 6 aborda os entregáveis após a elaboração do texto. Dessa forma, é citado sobre a implementação do observatório sismológico interativo e responsivo ao usuário final, bem como as telas das quais foram criadas no portal.

### 1.4.1 Cronograma

A especificação do cronograma de atividades, visa indicar quais serão os ciclos da pesquisa e quanto tempo levará para finalizar cada uma delas.

<b>Cronograma Pesquisa</b>	<b>Março</b>	<b>Abril</b>	<b>Maió</b>	<b>Junho</b>
Trabalhos relacionados	x			
Coleta dos dados		x		
Metodologia		x		
Início desenvolvimento			x	
Processamento dos dados			x	
Análise e interpretação			x	
Revisão do texto				x
Implementação do portal				x
Divulgação dos resultados para o público				x

Tabela 1: Cronograma de pesquisa

Vale ressaltar que o tempo necessário para finalizar cada ciclo do cronograma de atividades depende dos recursos materiais e humanos do qual os pesquisadores possuem. Portanto, a divulgação dos resultados obtidos nessa pesquisa está referenciado na tabela anterior (tabela 1), prevista a entrega na data do mês de Junho de 2022.

## 2 Fundamentação Teórica

Os fenômenos naturais que se originam por meio de tremores terrestres ocorrem desde o início do planeta. Desde então a humanidade sofria com as consequências de tais fenômenos, dos quais são capazes de mudar paisagens, clima, mortes e diversos outros fatores. Entretanto, os terremotos começaram a ser analisados cientificamente apenas após o terremoto que devastou Lisboa, em 1755. Considerado um dos terremotos mais fortes que atingiu a Europa, e segundo os sismólogos modernos, o tremor foi capaz de atingir uma magnitude de 9 na escala Richter, do qual gerou um tsunami e por fim tirou a vida de cerca de 100 mil pessoas. Uma das consequências desse forte terremoto foi o interesse da ciência sobre a sismologia, ciência da qual era pouco explorada até a época.

O primeiro grande avanço da sismologia como ciência ocorreu em 1880, com o desenvolvimento dos primeiros sismógrafos realmente eficazes pelos cientistas ingleses Gray, Milne e Ewing no Japão. Posteriormente, a Sociedade Britânica para o Desenvolvimento da Ciência criou um comitê para o estudo da sismologia, que acabou por permitir a instalação, sob a orientação de John Milne, de estações sismográficas distribuídas em uma escala global ([Eder C. Molina \[2019a\]](#)). Entretanto, atualmente, a referência mundial em relação ao monitoramento global de tremores terrestre acontece por meio do Serviço Geológico dos Estados Unidos, país do qual sofre bastante com terremotos, principalmente por conta da falha de San Andreas.

### 2.1 Conceitos sismológicos

A sismologia, é a ciência que estuda os sismos. Um sismo é um movimento brusco e repentino do terreno que se produz pela ação de forças no seio do globo terrestre. A sismologia tenta determinar as origens dos terremotos e o modo de propagação das ondas geradas por tais. Entretanto, entre os objetivos dessa ciência, destaca-se a prevenção para atenuar os efeitos destrutivos dos sismos bem como o monitoramento por meio de sismógrafos ou estações sísmicas.

Já o termo sismicidade se refere à descrição da atividade sísmica, do qual pode ser aplicado à Terra como um todo ou a uma região em particular. A sismicidade se refere, em parte, à geografia da atividade sísmica, ou seja, à distribuição dos terremotos em uma região ou no planeta inteiro. Além disso, o termo sismicidade se refere às características físicas da atividade sísmica: energia liberada, profundidade onde ocorre a liberação de energia, a frequência com que ocorrem os terremotos, a natureza dos movimentos crustais que causam os sismos e os efeitos macroscópicos produzidos ([Eder C. Molina \[2019b\]](#)).

#### 2.1.1 Magnitude

A magnitude é baseada em medições precisas da amplitude das ondas sísmicas nos sismogramas, para distâncias conhecidas entre o epicentro e a estação sísmica. Ela é expressa numa escala logarítmica, o que significa que o aumento de uma unidade da magnitude corresponde a um aumento de 10 da amplitude das ondas sísmicas que estiveram na base da sua determinação. Portanto, a magnitude está associada com a energia liberada pelo terremoto.

### 2.2 Intensidade

A intensidade sísmica de um terremoto não é definida a partir de nenhuma grandeza física que possa ser observada ou calculada. Intensidade sísmica classifica os terremotos pelos seus efeitos sobre objetos móveis e sobre as edificações do local, pela reação das pessoas no momento do terremoto e pelas marcas deixadas pelo terremoto sobre a paisagem local. A intensidade sísmica depende também da geologia local.

## 2.3 Placas tectônicas

Ao longo dos anos, novos métodos de estudo de como as forças internas e externas moldam a Terra, tem gerado abundantes novas informações e excitantes questões. Nas três últimas décadas do século XX, geólogos desenvolveram uma nova teoria unificadora que relaciona os processos dinâmicos da Terra aos movimentos de grandes placas que constituem a capa externa do planeta, teoria esta chamada de Tectônica de Placas. Esta teoria oferece um modelo abrangente para explicar como a Terra funciona (Gama).

A teoria da tectônica de placas considera a Litosfera como sendo composta por vários pedaços, que se encontram em movimento. Estes pedaços são denominados Placas Tectônicas. Atualmente consistem de 7 grandes placas e outras tantas menores. Elas se comportam como blocos rígidos que se movem por correntes de convecção existentes na astenosfera. As placas se movimentam de 3 a 11 cm por ano em diferentes direções e apresentam tipos de contatos distintos. Os contatos entre placas tectônicas são áreas extremamente instáveis da litosfera, aí se concentrando episódios vulcânicos e terremotos (Gama).



Figura 1: Placas tectônicas globo terrestre

Portanto, existem três tipos de movimento das placas tectônicas: convergente, divergente e transformante.

- **Movimento convergente:** Quando as placas se movem uma em direção à outra. O resultado desse tipo de contato são montanhas e cordilheiras.
- **Movimento divergente:** Quando as placas mergulham uma em cima da outra. O resultado desse tipo de movimento são profundas que atingem o manto, havendo a possibilidade da formação de vulcões.
- **Movimento transformante:** Quando as placas se movem unilateralmente, paralelamente entre si. É um tipo de movimento que causa bastante atrito e resulta em falhas nas bordas das placas, um exemplo seria a famosa San Andreas.

Então, os conceitos sobre placas tectônicas serão utilizados nas análises a seguir, servindo de insumo para a validação de hipóteses. Dessa forma, verifica-se como as 9 placas estão distribuídas no globo terrestre e suas respectivas bordas, bem como o movimento realizado de cada uma.

## 2.4 Escalas

De maneira breve, a geofísica possui duas principais escalas de medição para um sismo. Para medir a energia liberada pelo sismo, utilizamos a escala Richter, e para avaliar seus efeitos, utilizamos a escala Mercalli-Modificada.

### 2.4.1 Richter

De acordo com Célia Fernandes, geofísica e técnica em sismologia do IAG-USP<sup>3</sup>, a escala Richter foi desenvolvida em 1935 na Califórnia, Estados Unidos. Ela é calculada a partir dos sismogramas (registros dos sismos) produzidos pelos sismógrafos. Cada unidade de magnitude representa uma energia liberada dez vezes maior que o grau anterior, ou seja, um terremoto de 4 graus na escala Richter libera uma energia dez vezes maior que um terremoto de 3 graus.

### 2.4.2 Mercalli-Modificada

A escala Mercalli, que mede a intensidade dos terremotos, foi proposta pelo vulcanólogo italiano Giuseppe Mercalli em 1902, e alterada em 1931, quando passou a ser chamada de Mercalli-Modificada. Essa escala possui 12 graus, dos quais podem ser classificados como demonstrado abaixo.

- **I:** Não sentido.
- **II:** Sentido por pessoas em repouso ou em andares superiores.
- **III:** Vibração leve. Objetos pendurados balançam um pouco.
- **IV:** Vibração como a causada pela passagem de caminhões pesados. Chacoalhar de janelas e louças. Carros parados balançam.
- **V:** Sentido fora de casa. Acorda as pessoas. Objetos pequenos tombam e quadros nas paredes se movem.
- **VI:** Sentido por todos. Deslocamento de mobília. Louças e vidros se quebram. Queda de objetos. Rachadura no reboco de casas
- **VII:** Percebido por motoristas dirigindo. Dificuldade em manter-se em pé. Sinos tocam em igrejas, capelas etc. Danos, como quebra de chaminés, ornamentos arquitetônicos e mobília; queda de reboco; rachaduras em paredes, algumas casas podem até desabar.
- **VIII:** Motoristas de automóveis sentem o tremor. Galhos e troncos se quebram. Rachaduras em solo molhado. Destruição de torres de água elevadas, monumentos, casas de adobes. Danos severos a moderados em estruturas de tijolo, casas de madeira (quando não estão firmes com fundação), obras de irrigação e diques.
- **IX:** Solo rachado, como "crateras de areia". Desabamentos. Destruição de alvenaria de tijolo não armado. Danos severos a moderados em estruturas inadequadas de concreto armado e tubulações subterrâneas
- **X:** Desabamentos e solo rachado. Destruição de pontes, túneis e algumas estruturas de concreto armado. Danos severos a moderados de alvenarias, barragens e estradas de ferro
- **XI:** Distúrbios permanentes no solo
- **XII:** Danos quase totais

---

<sup>3</sup>IAG-USP: Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo



## 2.5 Medição e monitoramento

Como dito anteriormente, existem diversos artefatos ou locais propriamente para fazer-se o monitoramento e medição de atividades sísmicas. Dentre esses, podem ser citados os sismógrafos e as estações sismológicas.

### 2.5.1 Sismógrafos

Os sismógrafos são dispositivos que permitem avaliar os tremores que provocam os deslocamentos das placas tectônicas por meio de sensores, dos quais conseguem registrar as ondas sísmicas e tornar viável estudos sobre a propagação do tremor. Portanto, o sismógrafo, consiste num sensor que detecta e amplifica os movimentos do solo que por sua vez se chama sismômetro, e num registador que produz um registo visível do movimento, chamado sismograma.

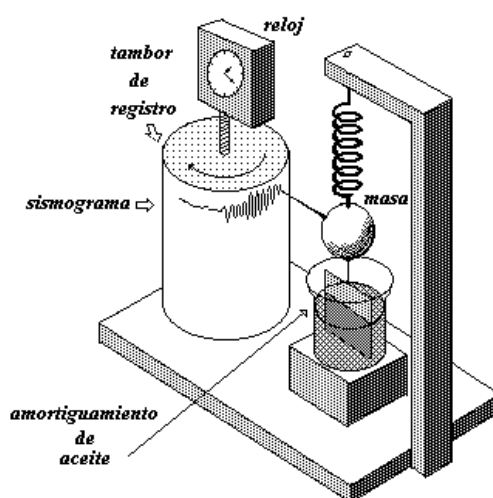


Figura 2: Sismógrafo Vertical

### 2.5.2 Sismograma

O sismograma representa a conversão do sinal do sismômetro para um registo temporal do evento sísmico. Nos primeiros tempos da sismologia moderna, o modo mais comum de obter diretamente um registo visível usava um tambor que rodava a velocidade constante de molde a providenciar uma escala temporal no registo. A invenção dos sismômetros electromagnéticos permitiu a conversão do sinal sísmico em sinal elétrico que é então registado.

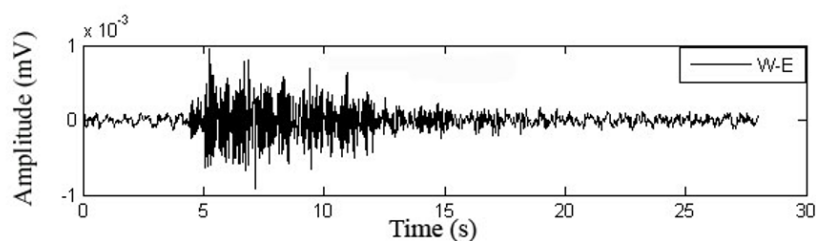


Figura 3: Exemplo de sismograma

## 2.6 Conceitos estatísticos

A principal vantagem de utilizar estatísticas em estudos de caso se desenvolve por meio do raciocínio analítico que enfoca a onipresença da variação, a investigação das fontes de variação, o planejamento de coleta de dados com a variação em mente, a quantificação da variação, e a explicação da variação.

Portanto, a estatística contribui para que dados gerem conhecimento e, como tal, deve ter como objetivo não só a produção de dados, como também a interpretação de dados já existentes, utilizando a combinação de gráficos, tabelas e medidas numéricas que permitam interpretar o que esses dados significam.

### 2.6.1 Medidas de tendência central

São medidas que tentam descrever o meio, ou a parte central, de um conjunto de dados. Algumas dessas medidas podem ser catalogadas como: média, mediana e moda.

### 2.6.2 Medidas de dispersão

As medidas de dispersão são medidas que representam a variação dos dados numa base de dados, algumas dessas medidas são catalogadas como a variância e o desvio padrão.

### 2.6.3 Formato e distribuição dos dados

O formato da distribuição dos dados tem como dever descrever como os dados são distribuídos, explicando então se a distribuição é simétrica ou assimétrica. Existindo três tipos elementares de distribuição de dados para uma variável. Ao analisar as medidas de tendência central, poderemos dizer em qual tipo de formato citado a variável analisada irá se identificar.

### 2.6.4 Modelagem e inferência

A etapa de modelagem dos dados tem como entregue a escolha e a construção de um modelo preditivo, do qual é avaliado através de testes de hipótese e conceitos estatísticos. Dessa forma, por inferência, o modelo visa responder à pergunta norteadora discutida inicialmente na etapa de entendimento do negócio, isto é, a previsão da magnitude de um terremoto dado as variáveis.

### 2.6.5 Árvores de decisão e floresta aleatória

Árvores de decisão são usadas para problemas de regressão e classificação. Uma das vantagens desse modelo é a facilidade visual como árvores, do qual origina o nome, e no caso de regressão, o início está como a raiz da árvore e seguem divisões com base em resultados variáveis até que um nó folha seja alcançado e o resultado alcançado.

Portanto, a floresta aleatória é um algoritmo de aprendizado supervisionado que usa o método de aprendizado em conjunto para regressão. O método de aprendizado conjunto é uma técnica que combina previsões de vários algoritmos de aprendizado de máquina (combinação de diversas árvores de decisão) para realizar uma previsão mais precisa do que um único modelo (Breiman [2001]).

Além disso, uma das principais vantagens na utilização desses modelos é a robustez em relação a dados discrepantes, isto é, outliers. Também, como explicado, o resultado é facilmente interpretado, principalmente por conta da visualização em ramos e nós que o modelo distribui suas decisões. Entretanto, deve-se considerar que a floresta aleatória é sensível ao overfitting. De maneira simples, overfitting é um termo usado em estatística para descrever quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

### 3 Revisão Bibliográfica

Segundo o estudo publicado por [Geller et al. \[1997\]](#) na revista Science, a hora, localização e a magnitude dos terremotos futuros não podem ser previstas de forma confiável e precisa de forma generalizada. O autor utiliza técnicas avançadas e atuais sobre a aplicação de métodos e sistemas não lineares para predição do fenômeno. Isso se dá, pois, um terremoto dificilmente ocorre de maneira isolada, realizando um efeito cascata, principalmente por condições no interior da crosta terrestre. Dessa forma, o autor cita intrinsecamente que tais fenômenos são imprevisíveis.

Da mesma forma, o texto elaborado por [Mondol \[2021\]](#) reflete sobre diversas abordagens e métodos para a predição de sismos na crosta terrestre. Para isso, foram testados diversos modelos estatísticos lineares e não lineares de modo a se chegar em um resultado para a previsão da magnitude de um tremor. No estudo foi possível verificar que a magnitude dos tremores referentes aos fenômenos de sismos difíceis para se realizar a previsão através de métodos convencionais.

Os fundamentos teóricos foram embasados em técnicas estatísticas de exploração, predição e inferência de dados. Por meio dessas é possível verificar a distribuição dos dados, dispersão e formato das curvas, além de ser possível fazer a escolha de modelos preditivos dos quais utilizam inferência estatística. Portanto, as técnicas citadas foram realizadas por livros publicados pelos autores [A \[2002\]](#) e [e BELFIORE Patrícia. \[2017\]](#).

O observatório sismológico demanda conhecimento preditivo para tentar inferir a magnitude de um sismo de acordo com respectivas variáveis. Então, em relação aos conhecimentos de modelagem de dados e predições, por métodos estatísticos e computacionais utilizaram-se referências, como estudos publicados por [Breiman \[2001\]](#) para aplicações das quais utilizam floresta aleatória, e estudos publicados em livros pelo autor [Aurélien \[2019\]](#).

Para um cientista de dados, um dos principais diferenciais é entender sobre a temática abordada no problema em que deseja-se resolver, nesse caso os terremotos, ou a sismologia. Para isso, decidiu-se utilizar materiais escritos pelos autores [Eder C. Molina \[2019a\]](#) e [Eder C. Molina \[2019b\]](#) para o melhor entendimento que agregasse em análises realizadas.

Então, tais textos dos quais abordam a mesma temática aplicada nesse estudo ajudaram nas tomadas de decisões em relações às análises e implementações do portal idealizado. Portanto, toda a criação desse artigo fez-se a verificação da grande parte dos pontos citados dos quais foram essenciais na implementação do observatório sismológico. Vale ressaltar que todas as análises realizadas consideram os materiais referenciados.

## 4 Metodologia

Em ciência de dados, a metodologia é uma estratégia geral que orienta os processos e atividades dentro de um determinado domínio. A metodologia não depende de tecnologias ou ferramentas específicas. Em vez disso, uma metodologia fornece aos cientistas de dados uma estrutura para proceder de acordo com os métodos, processos e heurísticas que serão usados para obter respostas ou resultados (Rollins [2015]).

### 4.1 Dicionário de dados

As informações a seguir foram coletadas por meio do Serviço Geológico dos Estados Unidos (USGS), do qual, como explicado anteriormente, será a fonte de dados para a confecção do observatório sismológico.

- **Magnitude:** Referente a magnitude do evento sismológico.
- **Tsunami:** Este sinalizador é definido como "1" para grandes eventos em regiões oceânicas e "0" caso contrário. A existência ou valor desta bandeira não indica se um tsunami realmente existiu ou existirá. Se o valor da bandeira for "1", o evento incluirá um link para o site da NOAA Tsunami para informações sobre o tsunami. O USGS não é responsável pelo alerta de tsunami.
- **Tipo:** Referente ao tipo de evento que gerou o sismo.
- **Alerta:** Referente ao alerta de impacto para determinado evento sísmico. Possui valores entre verde, laranja e vermelho.
- **Latitude:** Latitude em graus decimais. Valores negativos para latitudes meridionais.
- **Longitude:** Longitude em graus decimais. Valores negativos para longitudes ocidentais.
- **Tempo:** Referente ao horário que o evento ocorreu. Esse horário é disponibilizado em milissegundos.

### 4.2 Arquitetura da solução

Nessa seção é apresentado a arquitetura de solução utilizada para a criação do projeto. Então, por meio deste será diagramado quais os processos serão implementados para a construção do observatório, ferramentas utilizadas: consulta API, arquitetar a transformação e manipulação dos dados dos quais deseja-se coletar, e por fim a criação do portal de observações sismológicas.

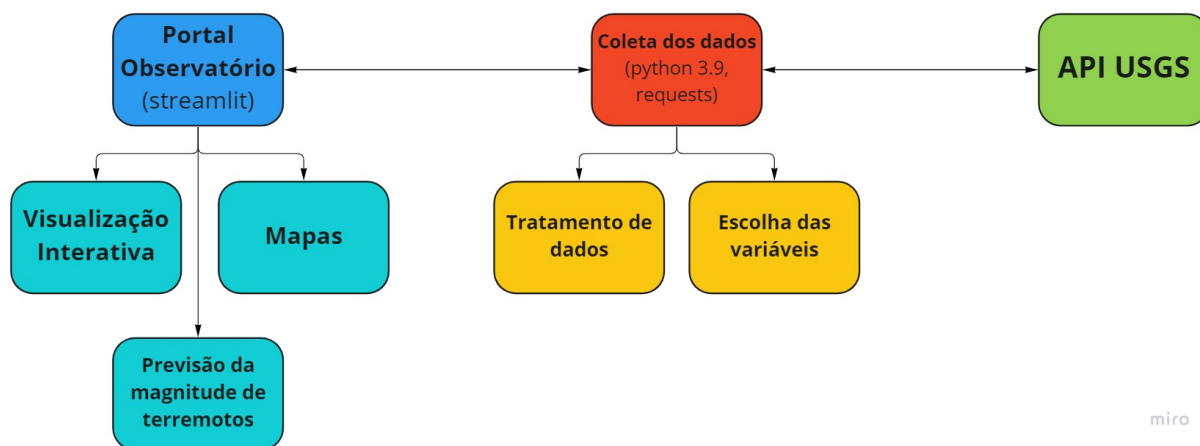


Figura 4: Arquitetura do projeto

### 4.3 Ferramentas utilizadas

A seguir serão listadas as ferramentas planejadas para a implementação do observatório sismológico. Portanto, abaixo estão listadas as linguagens de programação e suas respectivas bibliotecas, que servirão desde o processamento e tratamento dos dados até a utilização de modelos de previsão.

1. **Watson Studio:** Plataforma IBM para ciência de dados. A plataforma consiste em um espaço de trabalho que inclui várias ferramentas de colaboração e código aberto para uso em ciência de dados. A principal vantagem dessa ferramenta é a utilização dos recursos em nuvem para a extração dos dados presentes na API.
2. **Jupyter Notebook:** O Jupyter Notebook é um software de código aberto. Possui acesso online ou local, a ferramenta auxilia na compilação de códigos de diferentes linguagens de programação. É uma ótima ferramenta para documentação de códigos em markdown. Vale ressaltar que essa ferramenta será utilizada por meio da plataforma Watson Studio, ou seja, de maneira online pela nuvem IBM.
3. **Python 3.9 - Bibliotecas:**
  - **Pandas:** Manipulação, tratamento e análise de dados.
  - **Numpy:** Processamento de grandes dados, multi-dimensionais arranjos e matrizes, juntamente com uma grande coleção de funções matemáticas de alto nível
  - **Statsmodels:** Análises estatísticas.
  - **Scikit-learn:** Aprendizado de máquina, do qual torna possível, principalmente, a criação de modelos de predição.
  - **Datetime:** Manipulação de dados referentes a data e hora.
  - **JSON:** Tratamento de dados em formato JSON.
  - **Requests:** Requerimentos em APIs via Python.
  - **SQLAlchemy:** Mapeamento objeto-relacional SQL e conexão com o banco de dados escolhido.
  - **Plotly:** Criação de interfaces gráficas interativas.
  - **Stramlit:** Criação de interfaces web interativas.
4. **L<sup>A</sup>T<sub>E</sub>X:** Sistema de composição tipográfica de qualidade, que inclui recursos destinados à produção de documentos técnicos e científicos. Ou seja, é um sistema de preparação de documentos, que simplifica a digitação dos textos, especialmente de caracteres matemáticos

Portanto, destaca-se que o código final do qual será criado para guiar na confecção desse texto será disponibilizado no GitHub do autor, tornando o processo transparente e abrindo espaço para possíveis sugestões de melhorias para a solução do observatório sismológico.

## 5 Desenvolvimento

### 5.1 Extração e manipulação dos dados

Como explicado anteriormente, o observatório sismológico contará com duas páginas, do qual uma dessas irá conter a visualização dos terremotos em um mapa mundi. Para tal, abaixo é explicado de forma breve algumas manipulações que foram necessárias na etapa de coletar os dados disponíveis na API citada.

#### 5.1.1 Extração dos dados

A referência mundial em relação ao monitoramento global de tremores terrestre acontece por meio do Serviço Geológico dos Estados Unidos (USGS). Dessa forma, em seu site é disponibilizado uma API pública para a consulta dos dados, do qual pode ser acessada por diversas formas.

Esse estudo utiliza dados da API pública citada em formato JSON, coletando dados de tremores que tiveram magnitude maior ou igual a 2. Vale ressaltar que a API possui um limite de apenas 20.000 registros por requisição. Vale ressaltar que para pegar terremotos de maiores dimensões, configurou-se a API para coletar tremores com 4 graus ou mais. Consequentemente, será necessário fazer mais de uma consulta para coletar todos os dados do período citado na introdução.

Portanto, a etapa de extração e manipulação dos dados, será implementado com a utilização da linguagem Python em sua versão 3.9, para assim disponibilizar os dados para consumo do observatório sismológico. Como dito anteriormente, a etapa de inserção dos dados no banco de dados será efetuada diversas vezes, dado o limite de observações (20.000) por requisição da API.

#### 5.1.2 Tratamento dos dados

Um dos principais objetivos do observatório sismológico é o filtro por data. Dessa forma, percebe-se que a API disponibiliza os dados de data e hora de tremores em milissegundos. Portanto, foi necessária uma manipulação de tais dados para serem convertidos de milissegundos para o formato timestamp, ou seja, data e hora. Além disso, para facilitar a implementação do observatório, derivou-se uma variável referente ao ano do tremor ocorrido.

Já em relação ao tipo de terremoto, percebe-se que por motivos claros a API retorna tais observações na língua inglesa. Entretanto, para uma melhor compreensão do usuário final, mapeou-se tais palavras para que houvesse a tradução em língua portuguesa.

### 5.2 Análise exploratória

A análise exploratória dos dados visa entender e investigar os dados, principalmente numericamente, ou seja, como estão associados, as possíveis distribuições para assim extrair informações úteis nas próximas etapas do texto. Para a análise a seguir utilizaram-se 20.000 dados referentes a tremores nas datas de 15/12/2021 a 10/06/2022 filtrando apenas para terremotos de 2 graus na escala Richter ou mais. É válido ressaltar que toda a análise exploratória foi efetuada utilizando as bibliotecas Pandas e Seaborn na linguagem Python.

#### 5.2.1 Magnitude

A fim de verificar como os dados referentes a magnitude dos tremores se distribuí, fez-se necessário a aplicação de técnicas estatísticas para a avaliação, principalmente através de gráficos com as visualizações de tais registros.

Portanto, considerando que a existência de 20.000 registros de magnitude de tremores com 2 graus ou mais, dos quais são referentes ao período de 15/12/2021 a 10/06/2022, percebe-se o seguinte histograma com a distribuição de frequência desses dados.

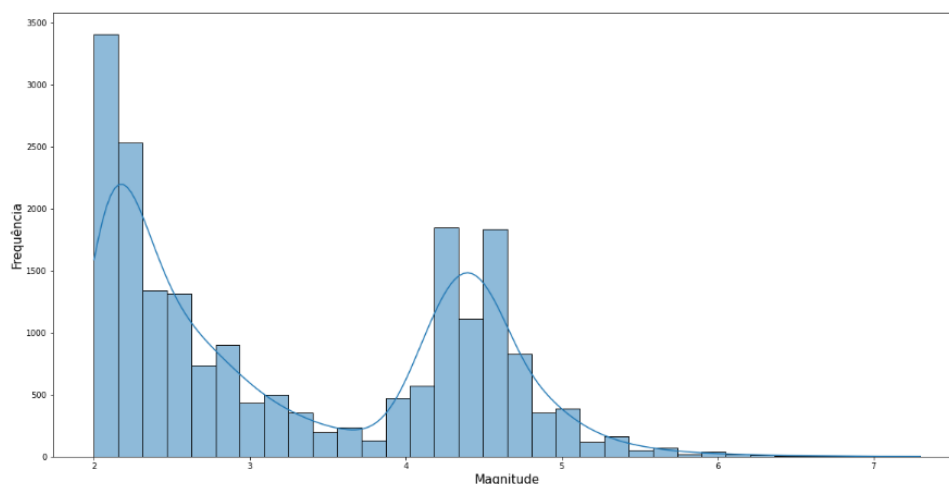


Figura 5: Magnitude — Histograma

De acordo com a distribuição do histograma referente aos dados de magnitude, percebe-se forte assimetria em tais dados. Esse comportamento era aguardado, uma vez que grande parte dos terremotos acontecem em bordas de placas tectônicas. Entretanto, entende-se que existem dois intervalos com grande frequência de tremores, o primeiro referente a terremotos com 2 a 2.5 graus na escala Richter e segundo referente aos tremores de 4 a 4.5 graus.

Já em relação às estatísticas no período coletado, os tremores tiveram uma média de 3.28 graus na escala Richter, com desvio padrão de 1.08 graus para mais ou para menos. Como demonstrado na figura 5, entende-se que 50% dos terremotos possuem magnitude menor ou igual a 2.9 graus. Por fim, nessa série histórica, o maior terremoto obteve 7.3 graus na escala Richter.

### 5.2.2 Local

Considerando a data pesquisada para a análise exploratória, procurou-se saber qual o local em que mais foram registrados tremores com 2 graus ou mais na escala Richter. Portanto, dada as possíveis possibilidades de localizações, decidiu-se encontrar as três principais localizações com maior frequência de terremotos, como é demonstrado abaixo.

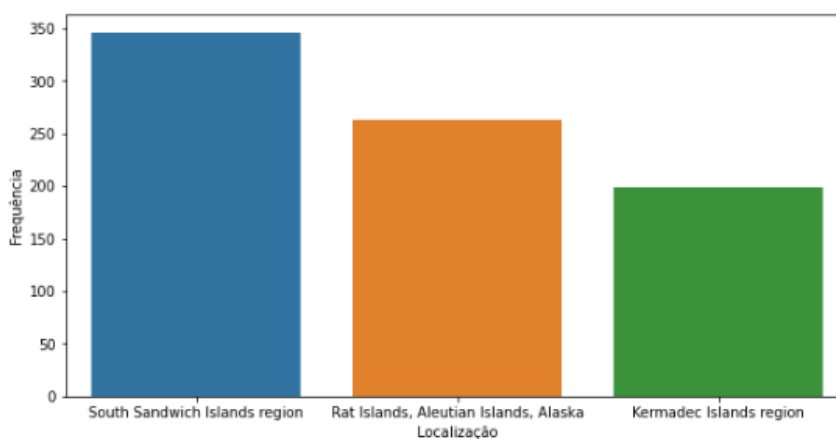


Figura 6: Top 3 localidades com maior frequência de terremotos

Portanto, a figura 6 demonstra que a localização de South Sandwich Islands Region, que em português são conhecidas como as Ilhas Geórgia do Sul e Sandwich do Sul, entre o continente sul-americano e antártica como a região com mais frequência de tremores, cerca de 346 terremotos no período de análise. As outras duas regiões são referentes ao estado do Alaska nos EUA e nas Ilhas Kermadec, território da Nova Zelândia.

Portanto, essa análise foca na associação de tais localidades às bordas das respectivas placas tectônicas. Através de tal é possível perceber que os terremotos, mesmo que em menor escala, acontecem de maneira mais frequente em regiões de bordas tectônicas. Tais dados possuem coesão com a realidade visto que, como explicado anteriormente, geralmente um terremoto é gerado a partir de uma ação entre duas placas tectônicas.

### 5.2.3 Tipo de terremotos

Em sua grande maioria, os tremores são gerados a partir de alguma atividade em bordas de placas tectônicas. Entretanto, quando se trata principalmente de terremotos com baixa escala, a atividade humana pode ser capturada por meio dos instrumentos sísmicos. Isto é, determinadas ações como explosões ou casos similares podem gerar pequenos sismos que serão identificados. Dessa forma, no período de 15/12/2021 a 10/06/2022 verificaram-se terremotos com as seguintes origens.

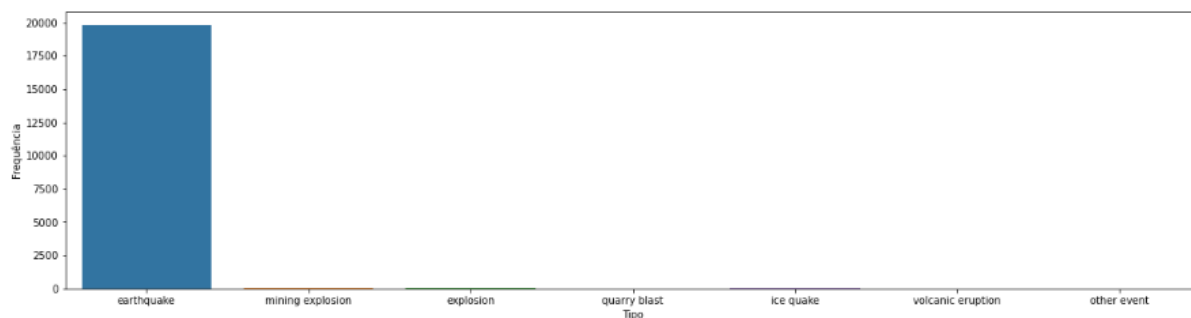


Figura 7: Tipos de tremores

A figura 7 apresenta o tipo dos tremores capturados no período de pesquisa. Percebe-se que majoritariamente foram terremotos oriundos de atividades tectônicas. Entretanto, ainda existem outros tipos, dos quais foram causados pela atividade humana. Para uma melhor visualização, abaixo estão listados os tremores referentes a causa humana.

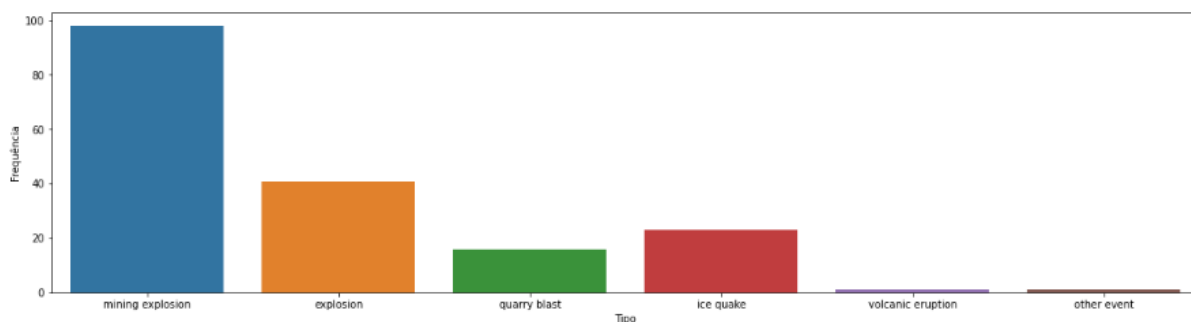


Figura 8: Tipos de terremotos (atividade humana)

Portanto, a fim de identificar as estatísticas em relação à escala de magnitude em cada um dos terremotos dos quais tiveram origem humana, idealizou-se a criação de uma tabela. Dessa forma, é possível identificar na tabela abaixo a magnitude média, mínima e máxima de cada tipo de terremoto causado por atividade humana. Para ajudar na criação da tabela, computou-se nesta os terremotos oriundos de erupções vulcânicas.



Tipo (EN)	Tipo (PT)	Magnitude média	Magnitude mínima	Magnitude máxima
Mining explosion	Explosão de mina	3.03	2.1	3.5
Explosion	Explosão	2.1	2	2.53
Quarry blast	Explosão de pedreira	2.19	2	3.17
Ice quake	Terremoto de gelo	2.3	2	2.5
Volcanic explosion	Erupção vulcânica	5.8	5.8	5.8
Other event	Outros eventos	2.2	2.2	2.2

Com auxílio da tabela anterior, verifica-se que as magnitudes médias, mínimas e máximas para cada tipo de terremoto. A maior média em relação à escala da magnitude fruto de atividades humanas, ou seja, exceto erupções vulcânicas, está em explosões de minas, com média de magnitude em 3.03 na escala Richter, constando determinada ocasião em que o tremor chegou a 3.5 graus. A prática de explosão em minas é utilizada em alguns casos, principalmente para a mineração em montanhas que possuem grande probabilidade de determinado mineral valioso.

#### 5.2.4 Significância

A significância de um tremor é referente a um número que descreve a importância do terremoto. De maneira simples, números maiores indicam um evento mais significativo. Esse valor é determinado por vários fatores, por exemplo: magnitude, MMI máximo, relatórios de feltro e impacto estimado. Dessa forma, abaixo é possível entender a distribuição das significâncias dos terremotos no período de pesquisa.

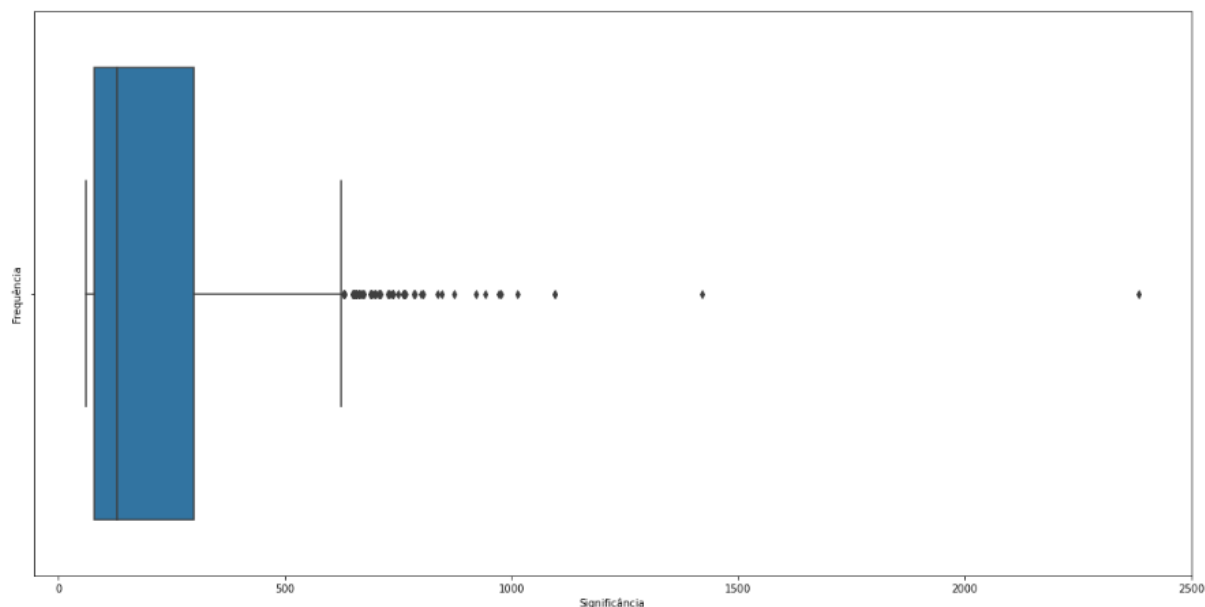


Figura 9: Significância — Boxplot

De acordo com o gráfico boxplot apresentado na figura 9, entende-se que a mediana da significância foi igual 129, isto é, até 50% dos tremores obtiveram significância menor ou igual a 129. É possível verificar diversos outliers, valores discrepantes em relação à distribuição. Portanto, tremores com significância acima de 626.5 são discrepantes. Observando a distribuição é perceptível um valor totalmente divergente, um tremor com significância igual a 2384, do qual é referente ao terremoto de 7.3 no Japão, tremor com maior magnitude na data pesquisada.

### 5.2.5 Profundidade

A profundidade de um tremor é referente a localização, em quilômetros, do seu epicentro, do qual é a zona central na superfície onde se manifesta. Vale ressaltar que, a escolha da profundidade de referência depende do método utilizado para localizar o terremoto, que varia conforme a rede sísmica. Como a API inclui dados de muitas redes sísmicas diferentes, o processo para determinar a profundidade difere para diferentes eventos. Dessa maneira, abaixo é possível verificar como se apresenta a distribuição das profundidades nos tremores pesquisados na data de 15/12/2021 a 10/06/2022.

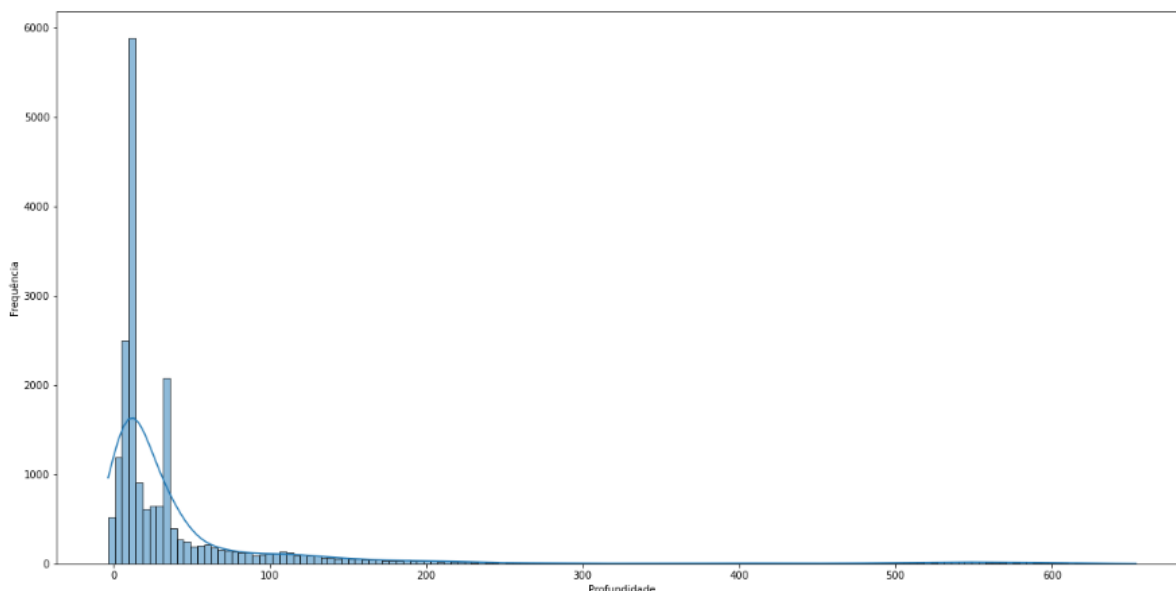


Figura 10: Profundidade — Histograma

Conforme a figura 10, percebe-se forte assimetria nos dados de profundidade. É possível identificar que a maior frequência da profundidade dos tremores são referentes a valores abaixo de 200 km. Em relação às estatísticas, a média de profundidade (em km) para os terremotos neste período foi de 47.77 km, com desvio padrão de 93.05 km para mais ou para menos. Até 50% dos tremores possuíam profundidade menor ou igual a 13.89 km. O terremoto com maior profundidade obteve o epicentro em 653 km, do qual é referente a um tremor na localidade de Fiji com magnitude de 4.3 graus na escala Richter.

### 5.2.6 Correlação

A correlação, da qual será usada nesse projeto será a correlação de Pearson, mais especificamente a matriz de correlação de Pearson. De maneira simples, a correlação de Pearson mede a força do relacionamento linear entre duas variáveis quantitativas. Interpretando a tal matriz de correlação infere-se da seguinte maneira tais intervalos: um intervalo de -1 a 1, quando igual a -1 a correlação é linearmente negativa, quando igual a 1 a correlação é linearmente positiva e se igual a 0, a correlação é nula.

Portanto, verifica-se a correlação das variáveis presentes no estudo do texto para entender a linearidade entre tais. Essa análise faz-se benéfica ao estudo, uma vez que poderá ser identificado quais métricas são os que mais impactam na magnitude final de um terremoto (além do fato de ocorrerem em bordas em placas tectônicas). Dessa forma, a matriz de correlação visa ajudar na identificação citada.

Além disso, a análise de correlação das variáveis é coerente com a possibilidade de implementação final do qual se aborda no texto, dado que a avaliação de linearidade e colinearidade das variáveis é benéfica ao estudo de implementação de regressão linear múltipla.

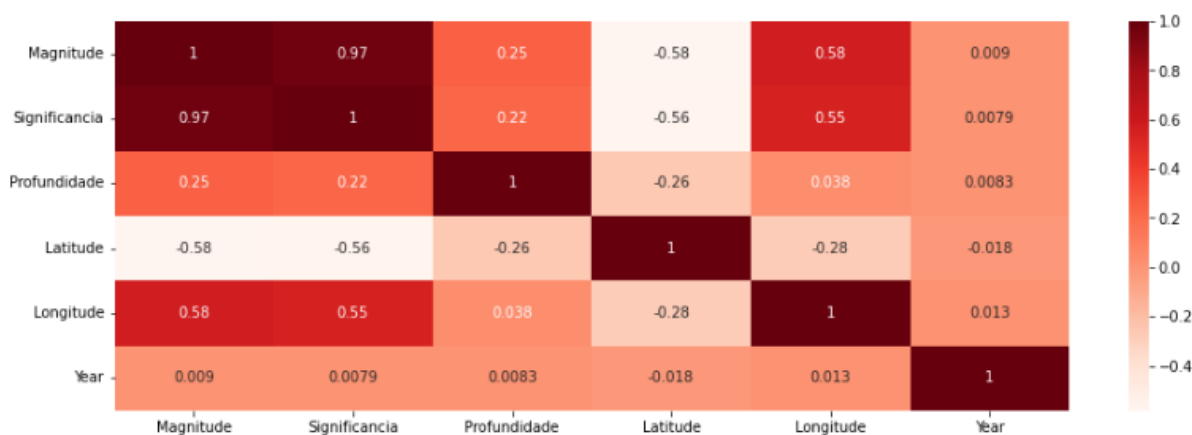


Figura 11: Variáveis numéricas — Correlação

A figura 11 representa a correlação das variáveis numéricas no período pesquisado. Quanto à variável alvo do estudo, magnitude, percebe-se que existem correlações positivas fortes, moderadas e fracas ligadas a tal. A significância possui correlação de 0.97 em relação à magnitude, isto é, uma forte linearidade entre as duas medidas, determinando uma possível colinearidade entre tais. A Longitude possui uma correlação igual a 0.58 com a magnitude no período pesquisado, implicando em uma correlação positiva moderada, cenário do qual já era esperado visto que tremores ocorrem nas bordas de placas tectônicas e as mesmas possuem grande extensão longitudinal. Por fim, a profundidade possui uma correlação igual a 0.25, configurando uma correlação positiva fraca.

### 5.3 Predição da magnitude — Floresta aleatória

O principal objetivo na criação de um modelo de predição da magnitude de terremotos em um portal, é aplicar variáveis que fazem sentido ao usuário final. Dessa forma, este artigo, a análise de regressão linear será utilizada para prever a magnitude em duas etapas, primeiro utilizando os dados de latitude e longitude fornecidos pelo usuário, e outra análise por meio da profundidade e longitude dos terremotos.

Como citado anteriormente no texto por Geller et al. [1997], terremotos são desastres praticamente impossíveis de se prever dada sua natureza incerta, o principal objetivo nessa etapa do estudo é entender o comportamento da previsão da magnitude de tremores através de algumas variáveis, para assim descobrir os melhores parâmetros.

Como explicado anteriormente no capítulo 2.6.5, a aplicação de regressão linear através de floresta aleatória possui fácil interpretabilidade e é robusta a outliers. Portanto, a decisão por tal modelo foi determinado principalmente pelos fatores explicados anteriormente, sobretudo após feita a análise exploratória nos dados.

Entretanto, segundo o texto publicado por Mondol [2021], onde há a tentativa de predição da magnitude de tremores a partir da latitude e longitude utilizando os modelos de floresta aleatória, Regressão Linear, Regressão Polinomial e LSTM, verificou-se um resultado não satisfatório, porém com a Regressão Polinomial desempenhando o melhor resultado para a predição da magnitude de tremores entre os modelos citados. Nesse estudo o autor propõe a variável profundidade do tremor para fazer a predição da magnitude dos terremotos, do qual se verifica uma boa relação com a magnitude no tópico anterior.

Portanto, o modelo de floresta aleatória será implementado por meio da biblioteca scikit-learn na linguagem Python. Vale ressaltar que, para encontrar os melhores valores para os hiperparâmetros, isto é, um parâmetro cujo valor é usado para controlar o processo de aprendizado do modelo, utilizou-se a função GridSearchCV, do qual faz o ajuste dos hiperparâmetros utilizando Cross-Validation. Então, o algoritmo utilizará 80% dos dados para treino e 20% para teste. Vale ressaltar que houve a normalização normal

nos dados, por conta das diferentes medidas numéricas para cada variável.

### 5.3.1 Latitude e Longitude

Visando uma implementação que faça sentido para a interação do usuário final no observatório sismológico, decidiu-se tentar efetuar a previsão da magnitude de um tremor, inicialmente considerando as variáveis de latitude e longitude informada pelo usuário. Entretanto, como citado no texto por Mondol [2021], tais variáveis não são tão eficazes na predição da magnitude de um terremoto. Por fim, fez-se a análise das métricas para verificar a viabilidade.

Como foi possível perceber na etapa de análise de correlação, a Latitude influencia negativamente na linearidade em relação à magnitude. Dessa forma, essa variável pode acabar atrapalhando o aprendizado do modelo. Com tudo, a medida de avaliação do modelo  $R^2$ , do qual demonstra quão explicativa a reta é em relação aos dados e possui variação entre 0 a 1, obteve um valor de 0.81. Entretanto, o que parecia um valor ótimo para a reta criada no modelo, verifica-se um overfitting, pois ao testar a latitude e longitude da capital do Brasil, local onde raramente ocorrem tremores, verificou-se uma previsão de terremotos com magnitude 5.37 na escala Richter. Abaixo está a representação dos primeiros nós referente a primeira árvore de decisão utilizada no algoritmo de floresta aleatória com as respectivas variáveis e hiperparâmetros utilizados.

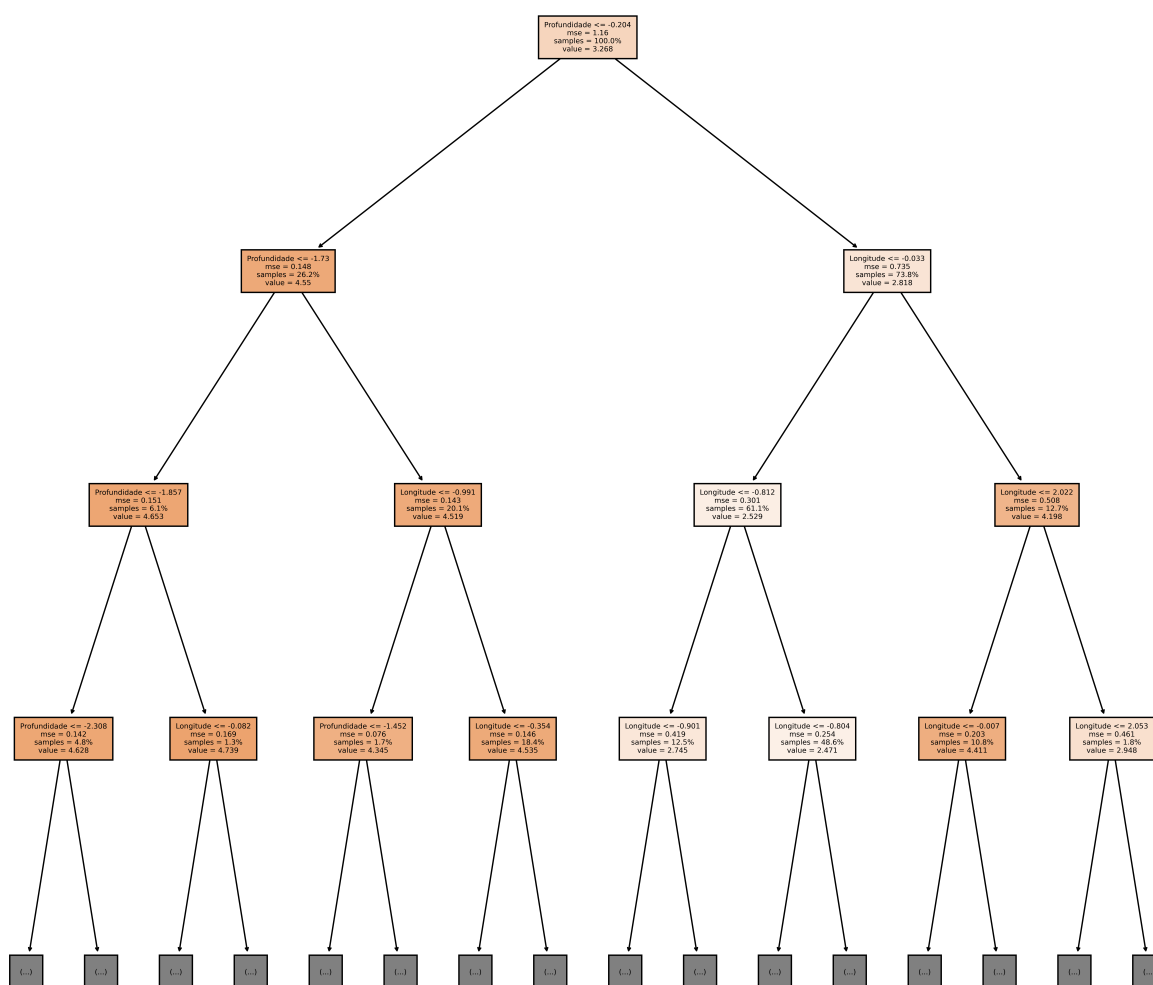


Figura 12: Exemplificação da primeira floresta com os nós iniciais

Deste modo, como demonstrado anteriormente, o modelo floresta aleatória obteve overfitting em relação à predição de magnitude de tremores dado a latitude e longitude. Porém, pensou-se em outra alternativa

visando os dados disponibilizados pelo usuário para ser possível a predição.

Portanto, apesar de bons resultados em relação ao modelo de floresta aleatória, verificaram-se predições não tão eficientes em regiões que não costumam ocorrer tremores. Vale ressaltar que, os terremotos são sentidos quando a magnitude está entre 3 e 4, também deve-se considerar a escala de Mercalli-Modificada, citada anteriormente.

### 5.3.2 Longitude e Profundidade

Como é citado no texto por Mondol [2021], a profundidade demonstra ser uma variável promissora na predição da magnitude de um tremor. Dessa forma, visando a disponibilização do modelo no observatório sismológico, decidiu-se fazer a análise do modelo de floresta aleatória utilizando a longitude e profundidade.

Considerando o cenário anterior, pensou-se na utilização das variáveis de longitude e profundidade disponibilizadas pelo usuário para prever a magnitude. Isto é, o usuário providenciará valores sobre tais variáveis para entender o seguinte caso: caso o epicentro de um terremoto tenha profundidade  $X_1$  (em km) e a longitude seja  $X_2$ , quão alta a magnitude deste tremor seria? Ao aplicar tal cenário no modelo floresta aleatória, verificou-se um  $R^2$  igual a 0.76. Ao testar utilizando a longitude referente a capital do Brasil, local onde raramente ocorrem tremores, e o epicentro com profundidade de 53 km, o modelo previu um tremor de 2.37 graus na escala Richter. Abaixo está a representação dos primeiros nós referente a primeira árvore de decisão utilizada no algoritmo de floresta aleatória com as respectivas variáveis e hiperparâmetros utilizados.

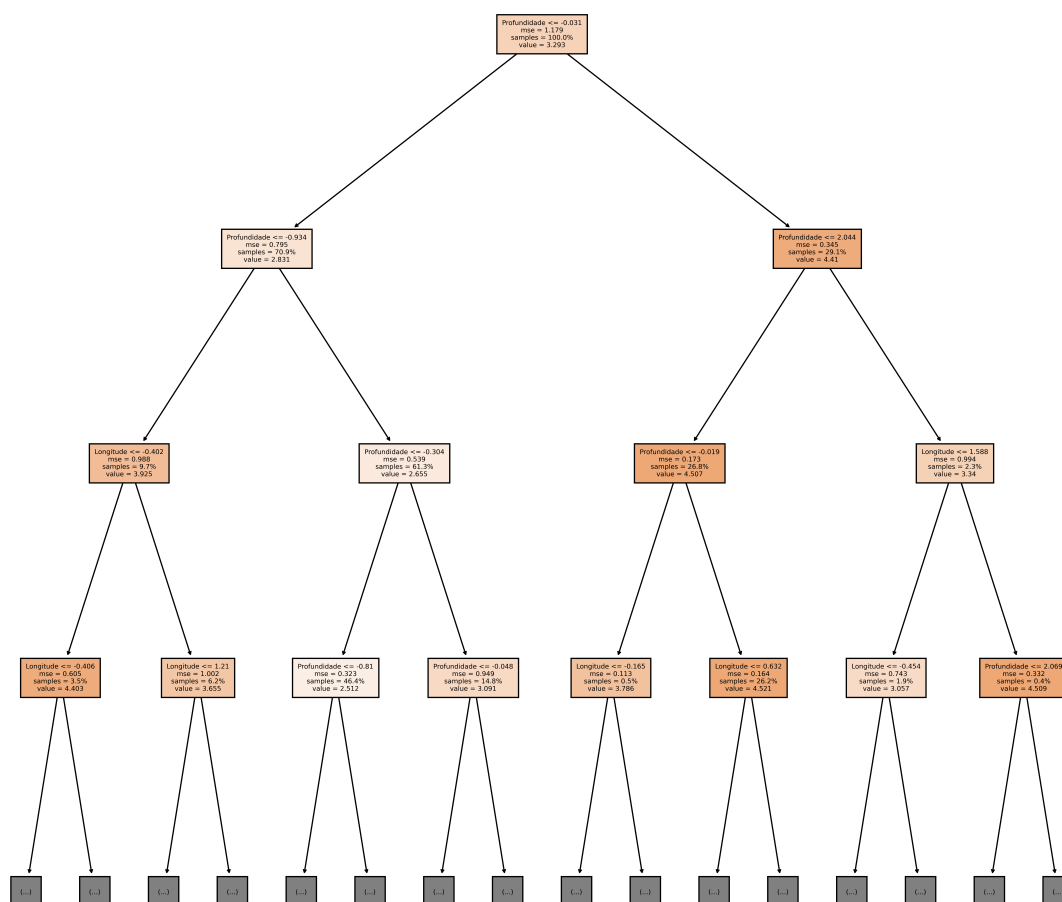


Figura 13: Exemplificação da primeira floresta com os nós iniciais

Portanto, como sugerido inicialmente, a utilização da profundidade apesar de diminuir a medida de  $R^2$ , traz bons resultados em relação aos testes feitos em localizações com menor probabilidade de ocorrência dos tremores. Então, decidiu-se aplicar as respectivas variáveis para realizar a previsão da magnitude de determinado sismo no observatório sismológico com os dados que serão disponibilizados pelo usuário.

## 5.4 Observatório sismológico

Um dos objetivos do observatório sismológico é trazer de forma fácil e clara aos usuários a distribuição dos tremores em determinada data. Entretanto, o portal será dividido em duas páginas.

- **Documentação:** Haverá uma breve explicação sobre o observatório sismológico, além de alguns conceitos que possam ajudar o usuário a entender melhor as respectivas visualizações.
- **Visualização:** Nessa página estará disponível ao usuário diversos filtros para que assim possa verificar os terremotos desejados. Decidiu-se por um componente interativo para que assim a visualização dos dados possa ser dinâmica.
- **Previsão:** Como citado anteriormente, um dos objetivos do portal é a interação do usuário, tornando possível a predição da magnitude de um terremoto conforme os dados disponibilizados para as variáveis escolhidas: longitude e profundidade.

### 5.4.1 Página inicial

Como citado anteriormente, a página inicial visa trazer alguns pontos referentes ao projeto implementado, bem como informações do autor e conceitos sismológicos para uma melhor experiência do usuário em relação à visualização dos tremores em escala mundial ou continental.

### 5.4.2 Visualização dos sismos

O observatório sismológico visa demonstrar os tremores ocorridos com os respectivos filtros preenchidos pelo usuário. Dessa forma, para a criação do portal, decidiu-se pelos seguintes filtros:

- **Magnitude** (O padrão de 4 graus na inicialização)
- **Data de início e fim** (O padrão de 2021/01/01 a 2022/06/10 na inicialização)
- **Tipo de tremor** (O padrão de Terremoto na inicialização)
- **Tipo de projeção** (O padrão de natural earth na inicialização)
- **Continente** (O padrão de mapa-múndi na inicialização)

Portanto, a ideia inicial dessa implementação era a idealização de um gráfico, em tempo de pesquisa, que consumisse dados da API e fosse capaz de gerar uma visualização similar à demonstrada abaixo.

Entretanto, como é observado na figura 14, os pontos dos quais representam os tremores em tal data não demonstram de maneira visual a respectiva magnitude. Portanto, decidiu-se fazer com que cada ponto seja proporcional à magnitude do terremoto visualizado.

Portanto, verifica-se que para a disponibilização da visualização dos dados no Observatório Sismológico deve ser implementado um gráfico que seja interativo e responsivo, para assim ser possível oferecer uma experiência melhor ao usuário final.

Vale ressaltar que, caso o filtro aplicado pelo usuário exceda os 20.000 dados possíveis de retorno por parte da API, deve-se avisar sobre tal situação indicando até qual data o último dado é referente.

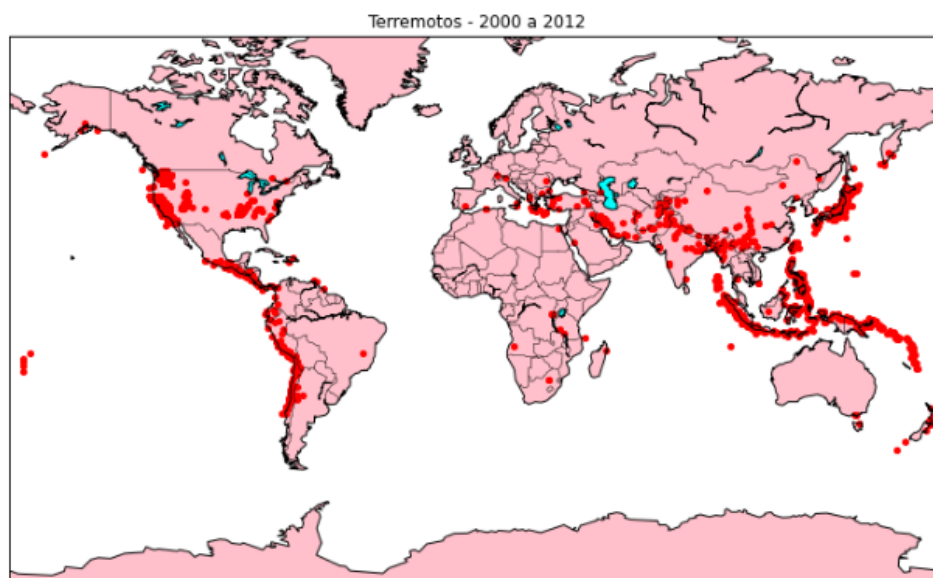


Figura 14: Idealização Observatório Sismológico

Dessa forma, visando encontrar uma biblioteca que atendesse as exigências citadas anteriormente, isto é, um gráfico em mapa que fosse interativo e dinâmico, utilizou-se a biblioteca *plotly* presente também na linguagem *python* para a implementação no Observatório Sismológico. Dessa forma, o usuário pode clicar nos terremotos, bem como saber as características de determinado tremor ao passar o mouse e outras diversas possibilidades. Abaixo é demonstrado a visão padrão ao acessar o portal, focando no círculo de fogo, bordas das placas tectônicas com maior número de terremotos.

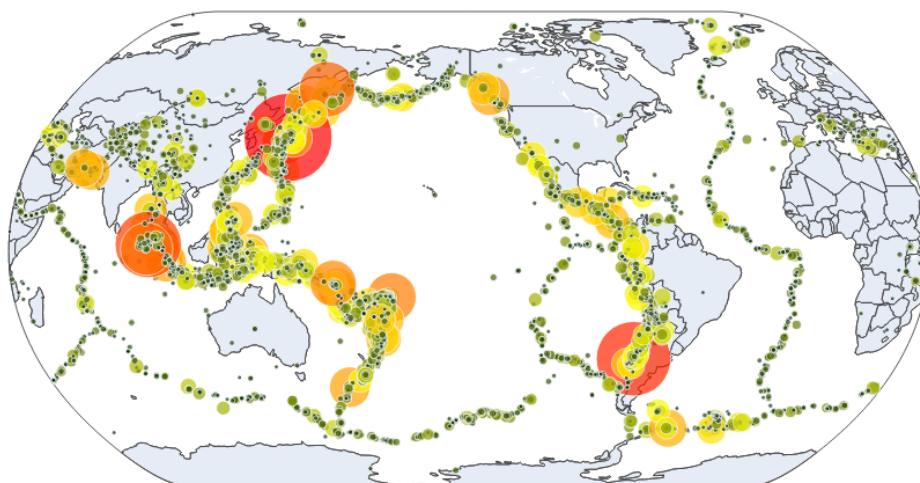


Figura 15: Mapa-múndi com os respectivos terremotos

Portanto, a visualização final da página da visualização dos dados possui o mapa com os respectivos terremotos na data de início e fim especificada, no lado esquerdo um simples menu com os possíveis filtros que podem ser aplicados. É importante citar que, caso o usuário faça algum filtro que não possua nenhum tipo de terremoto, o observatório foi configurado para não retornar o mapa e avisar através de uma mensagem de alerta sobre o resultado nulo oriundo da combinação de filtros feita anteriormente.

### 5.4.3 Previsão de sismos

Um dos objetivos finais e entregáveis dessa implementação é uma ferramenta onde o usuário conseguisse disponibilizar dados e então um modelo de predição fosse capaz de prever a magnitude do terremoto com tais características. Como demonstrado no capítulo 5.3, longitude e profundidade serão as variáveis necessárias para o modelo de floresta aleatória consiga realizar previsões da magnitude de um possível terremoto com tais características.

Portanto, haverá a funcionalidade de slicer visando barrar atitudes onde o usuário dê como entrada dados acima dos limites de profundidade e longitude. Consequentemente, os valores serão utilizados para fazer a previsão do sismo com as características abordadas pelo usuário, como é demonstrado abaixo.

## 5.5 Pontos de melhoria

Como citado anteriormente no projeto, a coleta dos dados é realizado por meio da requisição na API do serviço geológico dos Estado Unidos, da qual limita a 20.000 dados. Um dos principais pontos de melhoria seria a criação de um banco de dados que efetuasse a coleta histórica e diária dos dados. Essa abordagem melhoraria a velocidade de carregamento do portal, bem como poderia ser pesquisado um período maior sem a perda dos dados.

Na página de visualização, dado o limite de 20.000 dados por requisição na API, em alguns momentos dependendo o filtro não é possível verificar todos os dados daquele filtro, visto que há o limite citado. Dessa forma, uma melhoria seria a implementação de uma mensagem avisando sobre o limite do filtro e a data do último dado coletado.

Em relação à previsão, poderia ser estudado mais a fundo outras abordagens para melhorar o resultado predito. Etapas como limpeza e criação de novas variáveis a partir das existentes, seriam benéficos ao modelo de previsão.



## 6 Conclusão

O texto cita todas as etapas de arquitetura, análise e implementação de um portal que torne possível a visualização e predição dos dados referentes a terremotos oriundos da API do serviço geológico dos Estados Unidos. Da mesma forma, encontraram-se estudos com temas relacionados aos citados nesse estudo.

Para a criação do portal foi necessário avaliar os dados através de uma análise exploratória, do qual ajudou na etapa da escolha das variáveis para o modelo de previsão da magnitude. A análise exploratória demonstra a distribuição das variáveis numéricas, locais onde mais ocorreram terremotos em determinado período, e por fim os tipos e quantidade de terremotos causados por ação humana no período especificado.

Portanto, a criação de um portal para observações sismológicas visa ser interativo ao usuário que possui curiosidade sobre a temática. Para tal, criaram-se algumas telas das quais abordam assuntos diferentes. A implementação do portal realizada inteiramente em linguagem Python por meio de diversas bibliotecas possui código-fonte aberto ao público no repositório GitHub do autor. Por fim, o portal para a exibição da implementação pode ser acessado através do clicando aqui, .

## Referências

- BUSSAB W. O; MORETTIN P. A. *Estatística Básica*. Saraiva, São Paulo, 5 edition, 2002.
- GÉRON Aurélien. *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn TensorFlow*. Alta Books, Rio de Janeiro, 1 edition, 2019.
- Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- FÁVERO Luiz Paulo e BELFIORE Patrícia. *Manual de Análise de Dados – Estatística e Modelagem Multivariada com Excel, SPSS e Stata*. GEN LTC, Rio de Janeiro, 1 edition, 2017.
- Fernando B. Ribeiro Eder C. Molina. Sismologia tópico 01, 2019a.
- Fernando B. Ribeiro Eder C. Molina. Sismicidade tópico 02, 2019b.
- Jose Aparecido Gama. Unidade 1–introdução a geologia.
- Robert J Geller, David D Jackson, Yan Y Kagan, and Francesco Mulargia. Earthquakes cannot be predicted. *Science*, 275(5306):1616–1616, 1997.
- Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining and Machine Learning*. Morgan Kaufmann, Waltham, MA, USA, 3 edition, 2012.
- Manaswi Mondol. Analysis and prediction of earthquakes using different machine learning techniques. B.S. thesis, University of Twente, 2021.
- Akihisa Motoki, Luiz Carlos Chaves Novais, Kenji Freire Motoki, Leonardo Costa Oliveira, Adrienne Brito Lima, and Ricardo de Souza Fasolo. Observatório sismológico da universidade de Brasília. In *13th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 26–29 August 2013*, pages 1801–1806. Society of Exploration Geophysicists and Brazilian Geophysical Society, 2013.
- John B. Rollins. Metodologia de base para ciência de dados, 2015.