NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER ENGINEERING

INVESTIGATION AND FORECAST OF COMMON CATASTROPHE UTILIZING SPATIAL INFORMATION MINING PROCEDURE

M.Sc. THESIS

Naz Salah MUSTAFA

NICOSIA
April, 2022

Naz MUSTAFA

INVESTIGATION AND FORECAST OF
COMMON CATASTROPHE UTILIZING
SPATIAL INFORMATION MINING

MASTER THESIS

2022

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER ENGINEERING

INVESTIGATION AND FORECAST OF COMMON CATASTROPHE UTILIZING SPATIAL INFORMATION MINING PROCEDURE

M.Sc. THESIS

Naz Salah MUSTAFA

Supervisor
Assoc. Prof. Dr. Boran Şekeroğlu

NICOSIA
April, 2022

Approval

We certify that we have read the thesis submitted by Naz Salah Mustafa titled **"INVESTIGATION AND FORECAST OF COMMON CATASTROPHE UTILIZING SPATIAL INFORMATION MINING PROCEDURE"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of **Computer Engineering**

Examining Committee	Name-Surname	Signature
Head of the Committee:	Prof.Dr. Rahib Abiyev	....
Committee Member*:	Dr. Ahmet İlhan	....
Supervisor:	Assoc. Prof. Dr. Boran Şekeroğlu	....

Approved by the Head of the Department

...../...../20...

Prof.Dr. Rahib Abiyev

Head of Department

Approved by the Institute of Graduate Studies

NEAR EAST UNIVERSITY INSTITUTE OF GRADUATE STUDIES

...../...../20...

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

Declaration

I hereby declare that all information, documents, analysis, and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of the Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Naz Mustafa

Date: 25/04/2022

Acknowledgments

This thesis would not have been possible without the help, support, and patience of my principal supervisor; my deepest gratitude goes to Assoc. Prof. Dr. Boran Şekeroğlu for his constant encouragement and guidance. He has walked me through all the stages of the writing of my thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

Above all, my unlimited thanks and heartfelt love would be dedicated to my dearest family for their loyalty and their great confidence in me.

Naz Mustafa

Abstract

Investigation And Forecast of Common Catastrophe Utilizing Spatial Information Mining Procedure

Naz Mustafa

MA, Department of Computer Engineering

April, 2022, 78 pages

Natural disasters are unavoidable, and their occurrence has a significant impact on the economy, the environment, and human life. It assumes a principal part in dealing with the disaster level in a nation's repository. Devastation may occur as a result of building collapse, disease outbreaks, and natural catastrophes such as tsunamis, earthquakes, and wildfires, among other things. The world has the option to predict these catastrophes, particularly earthquakes, has never been so imperatively significant due to their natural occurrences. Spatial data mining is an innovative field when it comes to filtering out useful information from a load of raw data. The objective of this thesis is to design a model to analyze this catastrophe (earthquake) around the world using spatial data mining and give a predictable outcome using machine learning. This will reduce wreaking havoc on property, human lives, and critical infrastructure systems since the livelihood of a nation depend on it. Therefore, the aim of this thesis is to collect and apply the blend of spatial data mining and ML explicitly. These models will be evaluated and differentiated to perceive the best outcome. The datasets used for the experimentation consist of various earthquake occurrences around the world. It is believed that this work will develop an insightful model with high and strong accuracy to help people and specialists whenever thought about.

Keywords: Catastrophes, spatial data mining, machine learning, natural disaster

Özet

Investigation And Forecast of Common Catastrophe Utilizing Spatial Information Mining Procedure

Naz Mustafa

MA, Department of Computer Engineering

April, 2022, 78 pages

Doğal afetler kaçınılmazdır ve meydana gelmelerinin ekonomi, çevre ve insan yaşamı üzerinde önemli etkileri vardır. Bir ulusun deposundaki afet düzeyiyle başa çıkmada temel bir rol üstlenir. Yıkım, diğer şeylerin yanı sıra bina çökmesi, hastalık salgınları ve tsunamiler, depremler ve orman yangınları gibi doğal afetlerin bir sonucu olarak meydana gelebilir. Bu felaketleri, özellikle depremleri tahmin etme seçeneğine sahip olan dünya, doğal oluşumları nedeniyle hiç bu kadar önemli olmamıştı. Uzamsal veri madenciliği, bir dizi ham veriden faydalı bilgileri filtrelemek söz konusu olduğunda yenilikçi bir alandır. Bu tezin amacı, dünya çapında bu felaketi (deprem) mekansal veri madenciliği kullanarak analiz etmek için bir model tasarlamak ve makine öğrenimi kullanarak tahmin edilebilir bir sonuç vermektir. Bu, bir ulusun geçimi buna bağlı olduğundan, mülk, insan yaşamı ve kritik altyapı sistemlerine zarar veren hasarı azaltacaktır. Bunu yapmak için, tez, uzamsal veri madenciliği ve ML karışımını açıkça toplamayı ve uygulamayı planlamaktadır. Bu modeller değerlendirilecek ve en iyi sonucu algılamak için farklılaştırılacaktır. Deney için kullanılan veri kümeleri, dünyadaki çeşitli deprem oluşumlarından oluşmaktadır. Bu çalışmanın, düşünüldüğünde insanlara ve uzmanlara yardımcı olmak için yüksek ve güçlü doğrulukta anlayışlı bir model geliştireceğine inanılıyor.

Anahtar Kelimeler: Felaketler, mekansal veri madenciliği, makine öğrenimi, doğal afet

Table Of Contents

List of Figures

List of Tables

List of Abbreviations

SVM:	Support Vector Machine
CNN:	Convolutional Neural Network
KNN:	K-Nearest Neighbors
ML:	Machine Learning
ANN:	Artificial Neural Network
NN:	Neural Network
DL:	Deep Learning
SL:	Supervised Learning
DT:	Decision Tree
AI:	Artificial Intelligence
FIS:	Fuzzy Inference System

CHAPTER I

Introduction

Background of the Study

Natural disasters are unavoidable, and their occurrence has a significant impact on the economy, the environment, and human life. Devastation may occur as a result of building collapse, disease outbreaks, and natural catastrophes such as tsunamis, earthquakes, and wildfires, among other things. When earthquakes strike, millions of structures are destroyed as a result of the seismic impact (Mignan, 2019). Since the 1990s, an assortment of machine learning calculations has been used in the prediction of forest fires. In Italy, recent research used a spatial mining method to fathom the issue. The arbitrary timberland approach was utilized to outline the vulnerability of forests to fire in this research (Tonini et al., 2020). Floods are the most catastrophic natural catastrophes, wreaking havoc on property, human lives, and critical infrastructure systems. A set machine learning approach based on Random Forest (RF), irregular subspace, and support vector machine were utilized to map flood susceptibility (Islam et al., 2021). As the populace develops quickly, individuals require to arrive to live, and the result may be a loathsome unsettling influence on the biological system, causing worldwide warming and expanding the number of characteristic fiascos. People in less developed countries cannot withstand the damage to infrastructure caused by normal fiascos. The consequences of normal fiascos leave people dire and sometimes undetectable; moreover, rescue efforts are unable to be carried out in the majority of cases, and victims are unable to be recognized owing to topographical variables unique to each location. Catastrophes such as wildfires that spread swiftly in densely populated regions make firefighting more difficult; in such cases, it is critical to establish techniques to foresee such scenarios in order to avoid such disasters from occurring in the first place.

As technology continues to improve, aeronautical Systems have begun to apply smart technology to the development of unmanned aerial vehicles (UAVs) prepared with cameras, which can reach inaccessible ranges and photograph and videotape the effects of characteristic calamities on human life, establishment, and transmission lines. These pictures and recordings will be used to identify the consequences of characteristic catastrophes on human life, framework, and transmission lines. The data obtained from these drones make it possible to determine the victim's facial expressions, the level of a person's condition, and requirements in the aftermath of a catastrophe. It assists in taking action and carrying out the appropriate activities in the face of disastrous conditions. Raw photos captured by camera-equipped drones are analyzed, and an intensity analysis approach based on neural network feature extraction is used to extract information from the images.

A spatial mining method for two-dimensional cardiac attractive reverberation imaging reconstruction has been the procedure of acquiring imaging data to improve image analysis. The feature map of magnetic resonance images is reconstructed using a 10-fold technique using stacking of machine learning techniques. Thus, the feature extraction procedure becomes extremely rapid. Also, it takes not exactly 5-10 seconds to extricate the included framework from the feature matrix data. (Schlemper et al., 2017). Machine learning and spatial data mining provide a decisive technique, which is cumulative because of the coordinate input of multidimensional vector pictures, data processing, and recognition; these are the most commonly used technique for image interpretation and can be accomplished with a minimal level of difficulty (Tang, 2020).

Machine learning is of various categories and can be used to determine the scale of a normal calamity. ML strategies needn't bother with past information on the handling of data that gives it a benefit over different information handling frameworks (Wu et al., 2018; Hussain et al., 2021b). To investigate the exhibition of these spatial data mining, the co-connection coefficient will be a vital pointer in this thesis. The catastrophe expectation will likewise coordinate versatile ML methods for an expanded accuracy nature of the anticipated result. As a result of the above, this study will be carried out to compare and contrast alternative catastrophe prediction frameworks using geographical spatial data mining methods and view their accuracy using ML.

Statement of the problem

Disasters affect the lives and properties of people and animals worldwide. In many cases, the reason is not within our control. As pointed out by Brooks and Adger (2013), for three decades, namely 1970-80 (2nd place), 1980-90 (4th place), and 1990-00 (2nd place), India ranks among the five countries, leading in absolute numbers in loss of life. Not only is this the immediate effect seen in Dahar et al. (2013), but exposure to natural disasters in recent the chance of acute illnesses like diarrhea, fever, and other infections increasing in recent months. Acute respiratory diseases in children under five years of age, from 9 to 18% per year. It is directly related to the quantity and character of these consequences that a household's socioeconomic level is. Natural catastrophes have a significant impact on the home-based business as well. According to a report by Zheng et al. (2013), 40% of businesses that closed for three consecutive days went bankrupt or closed within 36 months. Disasters are also not uncommon. As for earthquakes (Nimmengadda, 2007), there are 20 earthquakes per year with a Richter magnitude more than or equal to 7.0. When calamity strikes, the consequences are significantly more severe in developing countries such as India, Burundi, Kenya, Rwanda, South Sudan, and other African countries.

As a result, an earthquake detection is an essential approach in catastrophe observing, catastrophe anticipation, and a variety of applications (Dong, 2009). Despite the fact that satellite imagery has a wide field of view and can cover huge regions, optical farther detecting pictures contain more content and can be viewed at a much higher determination than inaccessible detecting pictures. This is because optical remote sensing images have a better determination than inaccessible detecting pictures. As a result, optical imaging is believed to be more formally appropriate for documenting earthquakes than other methods (Yang et al., 2015). Spatial data mining involves replacing model building and specification with hierarchical characteristics (or classes) learning Recognizing latent properties of data from patterns in classes is a difficult task; it has shown considerable potential in detecting slip spots in object recognition scenarios that have been successful. This thesis will offer a spatial data mining process to analyze topographical catastrophes and their accuracy based on machine learning. When it comes to feature extraction and classification, it takes advantage of the compressed space of the crude picture. Feature Exploitation is a tedious problem in this respect. Finally, the characteristics are combined into a single categorization entry.

Aim and Objective of the study

The main objective of the study is to analyze catastrophes caused by earthquakes using spatial data mining techniques and consider its accuracy using ML. Specifically, the study aims to:

- Prepare an inventory of natural disasters (earthquakes).
- Assessment of the regulating factors of natural disasters (earthquakes) in the study area on the basis of analysis of previous earthquakes.
- Building and applying spatial data mining models of earthquake risk, assessing the applicability of different regulatory factors in the sensitivity assessment, and validating/comparing the results of the utilized ML technique.

Significance of the Study and Contribution

In the wake of recent research, many authorities and several stakeholders will benefit from its resources. Scholars and researchers around the world will also benefit from adding another key reference to their libraries. In particular, research will contribute to existing knowledge in the following ways:

- A complete overview of different data mining approaches that have been used in a variety of disaster-related activities.
- A detailed description of the many kinds and sources of data for each type of mission and catastrophe.
- A concise statement on disaster management "the current status" in the study context and presentation on recommended designing process to streamline disaster recovery.

Limitation

The limitations of this research work are as follows:

- In fact, land use, land cover, geology, and soil type have more impact on the occurrence of disasters (earthquakes), but the data available on a small scale and variability observed only on a large spatial scale, affecting the overall accuracy of the forecasts.
- Precipitation is one of the main contributors to earthquakes; however, rainfall data were not used in this study because data is scarce.

- All old earthquakes may not be considered when preparing a landslide inventory due to the unavailability of inventory. Therefore, it is based on images available on Google Earth, images from the world base, and a few field visits.

Proposal Organization

In terms of project organization, the thesis is divided into five chapters.

Chapter 1: Provides the background study, an overview, and a brief introduction to the research study. The statement of the problem and objectives, among others, are also provided in the same chapter.

Chapter 2: Presents Literature Review: the review includes various studies on the major subject of research. This chapter introduces the conceptual framework that forms the crux of the proposed research understudy—a review of extant works done by individuals in related fields.

Chapter 3: System Architecture: this includes a brief overview of the research methodology a and description of techniques used in the study

Chapter 4: To investigate the overall study findings and explain the dataset that was considered, the comes about, and discussions placed in Chapter 4.

Chapter 5: Finally, the proposed work is concluded in this chapter.

CHAPTER II

Literature Review

Concept of Natural Disaster

Natural failures are excessive activities inside the earth's framework (lithosphere, hydrosphere, biosphere, or climate) that range widely from the cruel, resulting in death or injury to people, to the unfortunate, resulting in the harm or misfortune of 'goods,' such as structures, to the fortunate, verbal exchange individuals, and harmed or hardship of `goods (Carter, 1991; Alexander, 1993).

The concern of catastrophe discount is a vital factor on the sector agenda. Nearly every day, failures seem inside the news's most important headlines. The number of recommended failures is increasing at an exponential rate, as is the number of suggested successes. Overall harm because of catastrophe, the guarantor misfortunes and the extent of casualties and those influenced. Homegrown disappointments disturb commercial center solidness, primary to soak decays in nationwide income by their compelling impact on the conveyance of number one commodities. Homegrown disappointments in numerous developing nations altogether restrict the financial boom. For example, the anticipated harm from Hurricane Mitch in Central America in 1998 was $6 billion, equal to 16% of GDP, 66% of exports, 96.5% of general constant capital, and 37.2% of general outside debt fairness for that year (Central American countries; 2000).

In 1989, the part states of the Joined Together Countries announced the 1990s the Worldwide Decade for Calamity Lessening (IDNDR) with the aim of "minimizing loss of life, property, and economic disruption." "The natural disaster poses a threat." IDNDR meets its purpose of assisting countries in drawing attention to the threat presented by natural disasters and the methods available to mitigate their effects. In the Yokohama Declaration of 1994, the United Nations said that it was referred to as 'publicity.' Innovation and information need to be made to be freed from fees and in a well-timed manner, specifically for growing countries. In 2000, the UN Countries framework was assigned as the suitable stage for implementing the International Strategy for Disaster Reduction (ISDR) as an extension of the IDNDR, aimed at engaging governments, international agencies, and the society. Typically driven by the significance of moving from a culture of danger reaction to a culture of hazard administration and anticipation. One of ISDR's goals is "to integrate existing hedging strategies into sustainable development plans through the cooperation of public, private and local communities through activities against works."

Effects of Disasters

In the period 1950-to 2000, natural disasters killed more than 1.4 million people, mainly due to earthquakes (47%), hurricanes (45%), and floods (7%). Economic losses estimated at over $960 billion are more or less equivalent to the occurrence rates of the most sorts of catastrophes: seismic tremors (35%), wind blasts (28%), and surges (30%). Other occasions (avalanches, dry seasons, fierce blazes, etc.) accounted for 7% of add up to misfortunes (Munich Re, 2000). Since 1950, there have been four times as many incidents as there were in the previous decade, and four times the number of claims have been filed as a result. It is additionally apparent that there has been a quick increment in the number of occasions. Several individuals back up this statement. Natural disasters are primarily related to catastrophes in several nations. Windstorms overwhelm several nations' expenses at ($90 billion), followed by seismic tremors ($25 billion). Back-up plans for unexpected disasters due to surges are altogether lower than (US$10 billion), as they are more extreme in developing nations, where protection thickness is lower (Berz, 1999).

The sharp increment in harm and individuals influenced by calamities is somewhat due to the improvement of communication, as almost no disasters receive media attention (Earth, 2000). However, it is also because of the increase in the world's population populace with common calamities. Several variables contribute to this, which can be broken down into those that lead to more prominent powerlessness and those that lead to a more noteworthy event of perilous occasions, such as:

- The world population is growing rapidly, from 3 billion in the 1960s to 6 billion in 2000, and the world populace is anticipated to be between 7 and 10 billion by 2050 (UNPD, 1999).

- Residing in areas already maintained a strategic distance due to affectability to normal risks.
- The concentration of populace and financial exercises in huge urban centers are generally found in powerless coastal ranges.
- The advancement of exceedingly delicate innovations and the expanding affectability of advanced mechanical social orders to framework disappointments. The recurrence of dangerous occasions related to climatic extremes (such as surges, dry seasons, violent winds, and avalanches) is expanding, which is connected to climate alter.

There's a converse relationship between the level of improvement and the misfortune of living in calamities. About 95% of casualties from natural disasters happen in developing nations, a region with a population of around 4,200 million people. Especially in industrialized countries, where caution frameworks are becoming more sophisticated, it is simpler to foresee the event of certain characteristic wonders and conduct Evacuations in large numbers. The implementation of restricted construction and zoning regulations is also a contributing factor to the decline in death rates in the industrialized countries.

Effect of Natural Disasters and Global Change

The impacts of climate alter are unmistakable and logical proof has been built up against numerous disturbing signs. The normal worldwide temperature has expanded over the past 100 a long time by almost 0.7 degrees Celsius, which is additionally outlined by an increment in sea temperatures. Agreeing to the Interval Board According to the Intergovernmental Panel on Climate Change (IPCC), the average global temperature will rise by 0.2 to 0.3 degrees Celsius every decade. In turn, this leads to exponentially larger dissipation rates and thus higher water vapor concentrations in the atmosphere, which causes large amounts of concentrated precipitation and tropical tornadoes. In numerous zones of the center and high scopes, there have been morphological changes within the dissemination of precipitation both by locale and by season. A noteworthy withdrawal of mountain ice sheets around the world can be watched, driving to higher levels of peril for frigid lake flooding and diminished water supplies. Dissolving ice caps, combined with the warm development of seawater, will cause ocean levels to rise. In common, the increment within the 20th century was more than 10 cm.

Estimated the expected increase in sea level over the next century is in the range of 10 to 90 cm, increasing the risk of coastal flooding in numerous ranges. Worldwide alter is additionally anticipated to extend wind storm action exterior the tropics. During the late 1980s and early 1990s, wind storms caused widespread damage in Europe (Munich Re, 2000). Warm waves ought to be more visited. The consideration comes around besides show up that El Nino and La Nina may as well be impacted by climate modification. Global warming has the potential to produce more ENSO miracles, as well as to heighten or draw them out. Many other changes have happened and will happen due to climate change. Land degradation, diminished biodiversity, and a surge in tropical illnesses outside of their natural transmission routes are just a few examples. There have been several attempts in later times to gauge the costs of anthropogenic climate alterations around the world. Within the long term, the take toll of anticipation methodologies is much lower than the harm caused by climate alter, which can be up to US$100 billion per year.

One way to bargain with common risks is to disregard them. In numerous parts of the world, not one or the other individuals nor specialists take calamity hazard genuinely, for a variety of reasons (socio-economic, political, cultural, religious). For effective disaster mitigation, a comprehensive disaster management strategy is required, also known as the fiasco administration cycle (UNDRO, 1991; Carter, 1991; Ingleton, 1999; Alexander, 1993). Fiasco administration includes two stages that take put sometime recently a calamity happens, fiasco readiness and readiness (the two combined are moreover known as fiasco relief), and three stages: post-disaster, disaster relief, recovery, and reconstruction. Unfortunately, in most countries, there is always a focus on disaster relief, and most disaster management organizations in the developing world have been established solely for this purpose. More recently, the focus has been on disaster mitigation, and in particular, on reducing vulnerability.

Investment firms, (international) funding offices, banks, and governments progressively require accurate information on the risks associated with hazards that could hinder investment or reduce the return on investment. Insurers and reinsurers also require a more detailed risk assessment to be able to set premiums for ventures. The standard process would also be to develop risk scenarios to minimize negative project consequences and financial loss. The ventures can be respectful development, lodging development, mining, rural and ranger service advancement, etc.

Geo-Spatial Requirements

Many forms of facts which can be wanted in herbal catastrophe control have each a critical spatial in addition to the transient component. Inaccessible detecting and GIS offer a historic hazard map that may be built from a database of known danger areas, highlighting which areas are unquestionably unsafe. Further detecting data must be coupled with a variety of data shapes, such as those derived from mapping, measuring systems, or testing centers, to develop characteristics that are favorable to the interior they observe at the time of disasters. Additionally, GIS may be used to adjust different risks and risks outcomes for the destiny change of a range of variables. The spatial modeling of perils could be a complicated task, wherein numerous components play a part, which may best be finished through specialists. It additionally entails a big wide variety of unsure elements, which ought to be taken into account. Hazard and hazard zoning has to be the idea of any catastrophe control assignment and has to offer organizers and choice producers whole and comprehensible data.

Inaccessible detecting information acquired from satellites is an effective device for mapping the spatial distribution of catastrophe-associated information over a brief duration of time. Many extraordinary satellite TV for PC structures exist today, with extraordinary traits associated with The spatial, temporal, and spectral resolutions of these instruments. There are many different sorts of natural catastrophes, such as floods, droughts, tornadoes, volcanic ejections, etc. can have beyond any doubt forerunners, genuine-time and close genuine-time adj for pc far off sensing might also additionally stumble on the early ranges of those activities as peculiarities in a time arrangement. When a catastrophe happens, the rate of records arrangement from discussing and region borne structures and the opportunity of records spread with a comparing quickness make it doable to uncover the predominance of the catastrophe. Simultaneously, GIS assessment can be utilized to plan departure courses, format centers for crisis operations, and integration fawning tv for pc data with distinctive, appropriate data.

Within the catastrophe alleviation stage, GIS is greatly valuable in combination with Worldwide Situating Frameworks (GPS) for looking and protecting operations. Farther detecting and GIS can help in harm appraisal and consequence observing, giving a quantitative base for alleviation operations. Amid the calamity recuperation stage, GIS can organize harm data and post-disaster examination data, as well as recreation destinations. Farther detecting upgrades, the database is utilized to recreate a range.

Catastrophe administration could be a multidisciplinary movement that requires spatial and worldly data and ability from numerous distinctive ranges of expertise, such as:

- Expertise in geographic information collection techniques, database creation, and disaster management information system design.
- Expertise in analyzing disaster phenomena, their location, frequency, scale, etc.
- Expertise in inventory items that could be destroyed if an event occurs: infrastructure, location, population, socio-economic data, and emergency relief resources, such as a hospital, firefighters, police officers, a police station, a warehouse, and so on. Organizations in emerging nations are putting spatial decision support systems, conflict management, and catastrophe management into practice.

Concept of Risk Assessment

Much of the exertion in calamity administration is political and social. In any case, choice producers require solid, up-to-date, and clarified data around the nature and geographic conveyance of risks and dangers, as well as conceivable chance scenarios. . Chance appraisal is considered the foremost central and critical perspective of calamity administration. Chance is characterized as “the anticipated number of passings, wounds or financial misfortunes from possibly harming occasions over a given period.” The primary fundamental prerequisite is to perform a quantitative hazard appraisal to get a quantitative chance outline. Most risk maps are still subjective and do not speak to “the likelihood of occasions with a certain degree of harm happening inside a given time frame.” In numerous creating nations, subjective hazard mapping is the, as it were, plausibility, given the restricted input information for the quantitative investigation. Information collection by inaccessible detecting and database plan for chance appraisal plays an imperative part, as do the Different sorts of modeling approaches are used based on the data and processes that are available. The tissue used for analysis, in particular, emphasis should be placed on the development of quantitative risk maps, which are developed by geoscientists based on probabilistic or deterministic models of the environment.

Another aspect that needs to be explored is the measurement of defenselessness, which is done by taking stock of the components at the chance (populace, building stock, fundamental hardware, transportation, and utilities). critical, high-loss equipment, operations) and an assessment of the extent of damage possible due to the event of a possibly harming wonder. Accentuation ought to be set on strategies for quick stock of at-risk things in thickly populated ranges (urban and country), utilizing high-resolution pictures and making databases of things at hazard, ought to be planned for numerous employments, based on cadastral databases. Another angle is defenselessness modeling, utilizing powerlessness bends in GIS. The contribution of the partners is also necessary to incorporate financial angles in arrange to supply a quantitative gauge of the harm. Combined danger and defenselessness data is utilized for quantitative hazard examination; counting adds up to misfortunes from diverse risks with diverse payback periods and levels. Strategies for information preparation and hazard evaluation are still generally to be created. Much of the ability is accessible in pro organizations (basically reinsurers that do not distribute their strategies for commercial reasons: e.g., MRQuake, MRStorm, and MRFlood from Munich Reinsurance (Munich Re., 2000).

At the worldwide The secretariat of the non-commercial level is the secretariat of the Worldwide Decade for Catastrophe Diminishment (IDNDR 19902000), Joined together Countries, Geneva, propelled the Span (Chance Evaluation Device for Diagnostics) activity urban ranges against seismic catastrophes) in 1996, with budgetary bolster originating from the Japanese government. It points to advance worldwide activities for seismic tremor calamity diminishment in urban regions, particularly in creating nations, by conducting considers within the nine cities examined. The three assigned universal establishing are GeoHazards Universal (GHI, USA), Worldwide Center for Calamity Moderation Building (INCEDE) / OYO Enterprise (Japan), and Bureau de Recherches Géologiques et Minières (BRGM, France), given specialized counsel to the cities examined through serious communication. Based on the involvement of nine cases considered, viable apparatuses for assessing seismic tremor harm and executing comparable ventures have been made so that all earthquake-prone execute comparable endeavors as a, to begin with, step in seismic chance administration. A comparative consideration to get the world's urban seismic chance was too carried out. More than 70 cities have taken an interest within the think about trade data. As partnered cities, more than 30 cities have joined Sweep to bring their profitable encounter to other cities (Span, 2001).

Whereas the Sweep strategy is still very common, a more complex strategy has been created within the Joined Together States called HAZUS. HAZUS could be a broadly appropriate standardized strategy actualized through a computer-based geographic data framework to assess potential harm from seismic tremors, surges, and winds. HAZUS was created by the Government Crisis Administration Organization (FEMA) beneath an organization assertion with the National Founded of Building Sciences (NIBS). HAZUS is right now able to evaluate seismic harm, and flood and wind models are being created (HAZUS, 2001). HAZUS could be an exceptionally advanced methodology that requires huge sums of information and isn't exceptionally attainable to actualize in most cities within the creating world within a sensible period.

In Australia, the Cities Project, or Urban Communities Geological Hazard Vulnerability National Project, as it was originally known, is a program of engineering development and applied research. It is designed to analyze and evaluate the dangers postured by a wide run of geographic dangers to urban communities (AGSO, 2001). One of the major challenges is the usage of these hazard maps in hazard scenarios and the advancement of spatial choice bolster frameworks for fiasco administration, which are utilized in:

- Anticipating the conceivable nature and degree of the crisis reaction required to reply to a catastrophe,
- Develop disaster recovery and reconstruction plans, and
- Mitigation Natural catastrophes may have a variety of outcomes.

The strategies center on application at the territorial and civil levels and, as a result, will be principal apparatuses in arranging and decision-making. A Catastrophe Administration Data Framework (DMIS) is subsequently required, which combines GIS with spatial choice back frameworks and is given through a client interface for low-cost GIS frameworks.

Previous Studies

The detection and recording of geological hazards on inaccessible detecting pictures have been examined in later a long time. Accomplishments in avalanche recognizable proof, catastrophe checking, and examination, early caution, etc., have been watched in Japan and a few European nations (Mantovani et al., 1996). To identify and recognize avalanches, most of the existing strategies utilize manufactured visual elucidation (Wang, 1999), object-oriented strategies, and factual models. The object-oriented approach utilized in Barlow et al. (2003) to distinguish avalanches in Landsat ETM+ pictures within the Cascades. It records avalanches utilizing advanced height show (DEM) information. A comparable strategy is utilized in Martin et al. (2005) for avalanche acknowledgment, which classified avalanches as landslide-dominant and landslide-dominant ground, accomplishing an exactness of 65 degrees.

The study of Nohani et al. (2019) compares four bivariate models; specifically, recurrence proportion (FR), Shannon entropy (SE), the weight of proving (Misfortune), and prove conviction work (EBF), in assessing landslide susceptibility assessment of the Klijanrestagh basin, in Iran has been realized. All methods give similar results, but the WoE model performs better among them. For the validation process, different techniques have been used by different authors. Usually, since the area under the curve is the main validation process, different techniques have been used to generate the AUC curve. Receiver performance characteristics in one of the main validation techniques are used. Similarly, other curves, such as the baud rate curve and prediction rate curve, are also used for validation (Chalkias et al., 2014). Research suggests that the area under the curve with a value greater than 70% is good for model validation (Khanh, 2009).

Khanh (2009) used the statistical index method to map landslide susceptibility in his study area. The landslide inventory is correlated with each calculation to calculate the ultimate avalanche affectability of the think about the region. The approval appeared to have an exactness that rivaled that of the calculated regression method, with an accuracy of 79% and 82% for the statistical information index method and the logistic regression method. One of the main advantages of using this method is that it allows researchers to calculate the weight of each layer separately. Using the logarithmic function to calculate weights for both positive and negative values helps to determine which layer has the most impact on landslides (Abidine & Abdelmansour, 2019). Pourghasemi et al. (2012) applied Shannon's entropy approach to analyze avalanche helplessness in Kalaleh town, utilizing eighteen avalanche conditioning components and the AUC evaluation results for accuracy. 82.15% confirmed. Similarly, Abidine & Abdelmansour (2019) also used the same method to map landslide susceptibility along with the frequency ratio method. A similar procedure mentioned earlier was utilized to create avalanche helplessness maps of the ponder region utilizing both strategies. The exactness of the show was confirmed by the AUC strategy, with FR giving a precision of 85.57% and SII appearing with a precision of 89.03%.

Bui et al. (2011) also used two methods for landslide susceptibility mapping, statistical index and logistic regression. Although some authors consider the above logistic regression method to be the one, the latter shows high comparability with other methods. The area under the curve of the logistic regression shows a value with 95% accuracy, and the statistical index method shows an accuracy of 94.02. Chalkias et al. (2014) also used statistical indexing. Since the model preparation method is the same, the AUC values were determined using the pass rate bend and the expected rate bend. It comes about to appear that the victory rate has an accuracy of 82%, while the prediction rate has 75%. The accuracy of this model depends on a larger number of landslide surveys. Although it is limited by other mapping techniques, it is one of the main techniques for landslide mapping with good model validation.

Related Works and Comparison of Techniques

Ponders analyzing the greatness of normal fiascos have gotten impressive consideration within the present ten-year period. A. Ashiquzzaman et al. (2020), a video source was used for fire detection; video feeds being processed could be an errand made conceivable by complex neural networks (CNNs). It necessitates high-level performance computing assets, counting illustration equipment, and, so, a location arrangement. Cost-effective and shrewdly fire is given on the premise of a complex neural organized design. In complex neural networks, a smoke detection model from forest fires called a densely expanded forest smoke network has been proposed by Li et al. (2019), which incorporates a candidate smoke zone division methodology utilizing a progressed organized engineering. Mangalathu et al. (2019) assessed earthquake-affected structures by investigating a profound learning strategy utilizing long-term memory.

Normal calamities are unforeseeable occasions, Hartawan et al. (2019) moved forward the multi-layer perceptron calculation by counting a complex neural arrangement executed on the raspberry pi to identify casualties of characteristic fiascos employing a gushing camera and offer assistance the group departure and protection of calamity casualties. Amit et al. (2017) proposed to apply programmed calamity location to a complex neural organize utilizing fiasco highlights from resized toady pictures of avalanche discoveries and surges. Ethereal pictures can appear in a more particular and bigger region of the ground, permitting a huge sum of data approximately the event of a catastrophe.

Social media systems, for example, Twitter, where individuals share their sees and data, have been utilized as information sources to perform catastrophe investigations. Yang et al. (2021), took earthquake-related data shared by clients on Twitter as a dataset and input it into a complex neural network-based real-time occasion discovery framework. The execution of the CNN module makes it conceivable to effectively identify a seismic tremor and inform the government of the seismic tremor in development utilizing instructive tweets. As tweets give a noteworthy sum of data, Madichetty et al. (2019) actualized a complex neural organize to carry out highlight extraction on both illuminating and ignorant tweets while categorizing the dataset, including tweets, with the assistance of a created neural network.

Social media is considered the most source of enormous information, with data being shared in the shape of pictures, recordings, and content; After a catastrophe strikes, social stages are full of distinctive sorts of data that offer assistance to responders to protect casualties. Much of the information contains unclear content, making it troublesome for protected groups to create precise choices. Nunavath et al. (2018) looked into past inquiries based on complex neural systems utilizing social systems as information sets and analyzed the adequacy of social media's enormous information in catastrophe administration.

An efficient highlight extraction technique has been linked to CohnKanade's amplified dataset to compare three items of the process, and this has been accomplished via the use of the two-layer architecture of a convolutional neural network (CNN). More than 90% execution rates, with more standard deviations, were accomplished by Boonsuk et al. (2019). The utilization of labor is troublesome in case of normal fiasco events in uneven ranges, and ceaseless electric control supply is profoundly influenced in these ranges due to upkeep issues of transmission lines. Hence, in this case, autopilot airborne hardware is utilized to assemble pictures, and covered-up substance from ethereal pictures must be distinguished in case of characteristic catastrophes such as avalanches and overwhelming snowfall. Zhou et al. (2019) evacuate clamor from crude airborne images and extricate catastrophe highlights employing a distinction procedure between pictures; They set up a complex neural organize to analyze the sort of calamity. In a few ranges, common calamities such as seismic tremors can happen due to topographical components. Finding the casualty in a brief time is exceptionally vital; Sulistijono et al. (2018) can perform airborne imaging.

Flooding could be a genuine and surprising calamity. Surges have an awesome effect on human life, influencing the economy and funds of nations. With the assistance of neural systems, it is conceivable to anticipate surges and spare the masses from fiascos. By executing complex neural systems and adjusted molecule swarm optimization (MPSO), Padmawar et al. (2019) created a profound learning approach to foresee surge circumstances and pre-identify people. Chen et al. (2019) proposed a forest fire detection image based on a drone image, stabilized the histogram, and applied a filter to smooth the images before examining them through a complex neural organize. Smoke location was performed employing a Nearby Parallel Demonstrate (LBP) and a Back Vector Machine (SVM). A comparison between handled pictures and the crude picture was made to check the adequacy of the proposed technique.

Wildfires greatly influence human life and financial circumstance, and finding casualties in a brief time could be a complex task. Convolutional neural systems are utilized to assist firefighters in finding casualties by recognizing smoke thickness from drone imagery. Simple CNN-based feature extraction with a proposed approach based on AlexNet Unique Deconvolution (SFEwANSD) allows the development of a real-time fire checking framework Gonzalez et al. (2017) and Samudre et al. (2019), effectively made strides in reaction times, decreased control utilization, and optimized execution by utilizing CNN's network-layer pipeline, running on field-programmable portal clusters. Due to the low spatial resolution of satellite images, these images cannot be used to detect forest fires; Lee et al. (2017) modified a deep complex network for high spatial resolution images, VGG13 and Google Net, using drones, disaster forecasting systems, web visualization systems, systems warning, and disaster response scenarios database and got very accurate results right from the start of forest fire detection. Catastrophe administration organization is active work to evaluate harm caused by calamity. Utilizing pictures gotten from social media amid and after four major common calamities, Nguyen et al. (2017) proposed a strategy of tuning CNN's work to particular occasions and cross occasions.

Direkoglu et al. (2020) propose a strategy of creating movement data pictures by computing optical stream vectors and utilizing CNNs; The proposed method effectively distinguishes between normal and abnormal human behavior in the event of a fiasco. UMN and PETS2009 datasets were utilized to perform the tests. Yuan et al. (2019) proposed Waveform Neural Organize (WNet) to name smoke thickness in pictures, which could be a troublesome errand; that is why a virtual dataset is created. The legitimate encoder-decoder architecture was collected to maximize input for extricating data from smoke thickness pictures and proposed WNet. The exactness of the suggested framework is progressed by giving the already encoded yields to the interpreting layers and combining them. A few information mining applications have been actualized utilizing social media substance; User-generated substance makes a difference amid disastrous occasions picking up a huge sum of data. The CNN show is utilized to extricate the surge picture from the crude picture, and color channels are utilized to refine the required location. Within the work of Layek et al. (2019), the proficiency and precision of the suggestion framework have been tried on many datasets and outflanked other strategies to deliver the leading comes about. The multi-layer agglomeration neural organize proposed in this think is utilized to distinguish and classify common calamities, as clarified within the strategy area. Moreover, compare a few of the progressed strategies displayed in Table 1.

Table 1.

Comparison of the techniques

Reference	Methodology Name	Outcomes	Weakness
Amezquita-Sachez et al., (2017)	Signal processing, image processing, and statistical method are all examples of signal processing.	Natural catastrophes should be predicted with greater precision.	Constrained measurable parameters for prediction
Zhang et al., (2014)	Particle swarm optimization	Anticipate the size of seismic tremor	Work as it were for expectation on the seismic dataset
Adeli & Panakkat (2009)	Neural network	Anticipate the greatness of seismic tremor	Restricted parameters used for prediction
Kradolfer, (2013)	Text mining, regular log mining technique	Seismological data may be used to quickly and accurately detect earthquakes.	Depends on public feedback to detect earthquake
Merz et al., (2013)	Decision tree Artificial neural network, genetic	Utilize a few parameters to get to the model for surge harm region detection Sum-up great comes about as compared to	The identification of surge-damaging zones is limited by a set of parameters. Storm surge preparations will be carried out in June and July.
Sahay & Srivastava (2014)	algorithm and wavelet transfer technique	as of now existing procedures in Southeast Asia	For time-series data, September is the best month to look at particular areas in India.
Venkatesan et al., (2012)	Back vector machine and naive Bayes are two types of algorithms.	Classify the normal fiascos on various parameters	Restricted for as it were early stages of natural disasters

Concluding remarks

The concept of natural disaster, its classification, terms and terminologies, its relation with different geo-environmental factors, mechanism, and related studies behind the assessment methods are thoroughly reviewed in this chapter. The necessity for the natural disaster assessment, general information, and geological description of the study area will be provided shortly. Literature showed several approaches that can be used for susceptibility assessment. Among which FR, SII, and SE are deterministic bivariate statistical methods that have provided satisfactory prediction results under a similar type of study area and causative factors in the previous studies. The methodological steps for the overall assessment will be described in the Methodology section.

CHAPTER III

Methodology

Here, the research topic is introduced, as well as the data sources and methods used.

Area of Study and Data Used

Earthquakes occur in many places of the world, including the United States, Bulgaria, and Turkey, to name a few. Because of the readily available data on the internet and its proximity to the study, this study chose to focus on earthquakes around the world. Karahacılı and its environs, notably Emirler, İnsu, Değirmençay, Uzunkaş, and Turunçlu, are noted for their frequent earthquakes. The samples for the study were taken from the records available after severe quakes from the 90s to the 20s, respectively. The settlement of Karahacılı has a latitude and longitude of 36°.813081N and 34°.482932E, a total area of 786.03 ha. There were collapses of energy transmission lines, as well as major transit roads and agricultural farms damaged as a result of the earthquakes that happened in the region. In the course of this inquiry, it was discovered that it affected the region tremendously.

The study made use of the earthquake occurrence over the globe on the world map, as illustrated in Figure 1 below. The earthquake site was chosen from a significantly impacted region of clayey limestone from the Langhian and Serravallian. Then, using a 1:250,000 scale geology map, the complete Lower-Miocene a geological formation called the Kaplankaya Formation, which covered the whole region, was discovered in the area scanned. Non-cohesive material consisting of sand and gravel was also found in the earthquake-affected regions and lime-stone. During the data collection, no active fault lines were discovered in the study area's local proximity. However, debris was discovered in the earthquake-affected region, which raises the likelihood of earthquakes. Previous investigations in the area revealed that the research area was located near earthquake zones that were currently active. (MTA, 2021; Can et al., 2009). Due to severe earth crust, the rubble behind the area's crown triggered an earthquake. The characteristics of the dataset are given in Table 2, and details of its depictions are given in the resulting section, as the data will be spoken to visually.

Figure 1

World map of the study area

Table 2.

Characters and meaning of datasets

Characters	Meaning
Area	Catastrophes (Earthquake)
Associated task	Spatial mining/ML
Data characteristics	GIS
Dataset No	10000
Missing value	N/A
Dataset format	CSV
Outcome	Features extraction and prediction.

Materials used for the study

In the course of this study, the following materials were used, and it's depicted in Table 3.

Table 3.
The information and characteristics that were utilized in the research

Name of the data set	Earthquake occurrence around the world
2020 Mosaic Orthophoto	35 cm, WGS84
Model of the Digital Elevation (DEM)	10 m, WGS84
2019 UAV Orthophoto	Approximately 2 cm,
2020 Sentinel	15 m, WGS84
October 2021, December 2019	WGS84
Google Earth images	Shape data
Lithology map	1:100,000, WGS84
Boundary of Mersin	ITRF 94
Meteorology data	2019–2021, Daily earth vibration

Earthquakes are natural occurrences caused by an identifiable trigger element such as torrential earth vibration (Hao et al., 2020), snowmelts, earthquakes, and volcanic causes, which are recorded in the event earthquake inventory (Legorreta et al., 2014). Throughout January and February, an earthquake occurred in Karahacli, and March 2019 was reportedly triggered by excessive earth vibration. Earth vibration data was collected using the Automatic Weather Observing System, which was located a few kilometers from the earthquake location. Graphic and polynomial trend lines were used to display 24-month earth vibration data for 2019 and 2020, with R^2 values of 0.74 and 0.82, respectively. In December of this year, the area got more than 400 kg/m^2 of precipitation. The precipitation rose by more than two times in January 2020, the preceding year in terms of comparison

A 2019 orthomosaic map with an orthophoto collected by UAV in 2020 was used to build an earthquake inventory map for this investigation. As in previous experiments, the flight, which was carried out using Anafi Parrot on June 28, 2020, was used as a determinate, with 70 percent frontal overlap and 80 percent side overlap, with a geometric standard deviation of 1.86 cm (Alptekin et al., 2019). A total of 447 photographs were picked and used from the trip, spanning a 21.1-hectare area. It used a 36-3° projection mechanism called TUREF (ITRF96). The study area accounted for around 10.3 hectares of the 21.1 ha flying area (Kusak et al., 2019). This research used there are a total of five GCPs (GCPs). However, even though GCPs used When the points were set up to show a uniform distribution, full homogeneity could not be achieved because of the large level difference in the slope, the loss of the ability to descend to the stream bank, given the fact that the summit of the earthquake region was perilous and sliding was still going on. For the GCPs, Continuously Operating Reference Stations were used to calculate the coordinates and elevations. This measurement has inaccuracies of 2.37 millimeters in each of the three dimensions: XYZ.

Earthquake threat assessment relies heavily on topographic factors. It is determining the greatest by utilizing a DEM to calculate the slope angle (Fenton et al., 2013). Maintaining shallow-surfaced areas with clayey deposits and a slope component was found to be useful in earthquake situations development in a study conducted by Capitani et al. (2013). As one of the hydrographic factors, drainage systems contribute to the creation of earthquakes by saturating the bottom and upper surfaces with water. Maps of slopes, aspects, and streams were created as a consequence and constructed using 5m key components such as topography and hydrography before an earthquake may be assessed using high-resolution DEM data.

NDVI was employed as a supplemental preliminary factor in this study. The NDVI (Normalized Difference Vegetation Index) ranges from +1.0 to -1.0. The NDVI is a dimensionless index that measures the difference in land cover reflectance between visible and near-infrared light and is used to measure the density of green on a piece of land (Weier and Herring, 2000). This can also be used to measure the density of damage caused by the quake on a piece of land. In the field of vegetation analysis, the well-known normalized ratio between red (R) and near-infrared (NIR) is the well-known NDVI, which is one of the most often used vegetation indicators indicated in the Equation. 1.

$$NDVI = \frac{(NIR \quad)}{(NIR+R)} \tag{1}$$

A healthy photosynthetic vegetation index runs from -1 to 1, with higher values suggesting healthier photosynthetic vegetation and lower values indicating stressed or non-photosynthetic vegetation (bare soil). To evaluate how the forest is faring, the NDVI used plant cover of earthquake regions has changed (Khan et al., 2019). The picture from Sentinel-2 taken on December 25, 2019, was utilized in this study.

Procedures

The methodologies employed in this study were adapted from Lutfiye et al. (2021), as explained in the sub-sections that follow. First and foremost, information on the inventory maps was provided. Furthermore, approaches to spatial data mining were used and reviewed, as well as the formulae used and explained. Figure 2 shows the procedure or the proposed process.

Under Appraisal, a couple of information preprocessing tasks referenced underneath were performed to improve on the handling of the tremendous and changed nature of the informational collection.

Information Reduction: Since the accessible informational index has plenty of cases, an information decrease was performed. A decrease of dimensionality was executed by thinking about something like 5 of the absolute 12 attributes accessible. Any remaining repetitive and unessential traits were disposed of from the informational index.

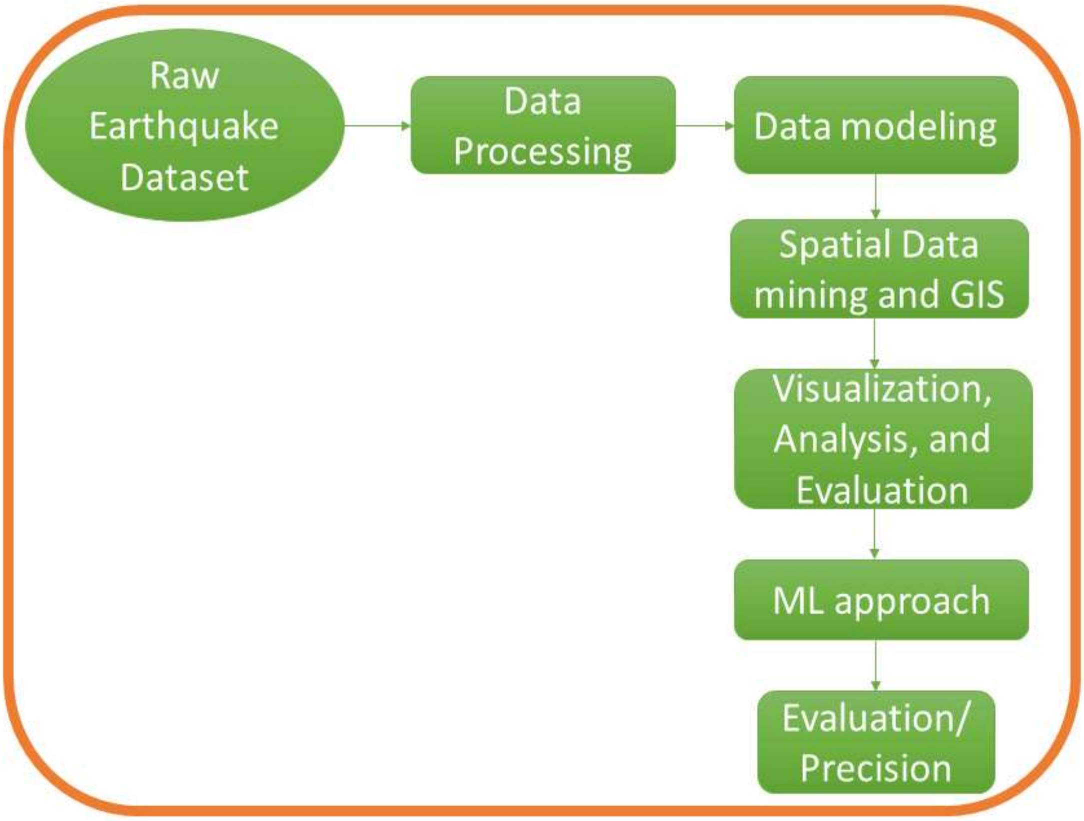
Information Integration: Credits Time Slot and Time Occurred; Area Name and Area ID were bound together under ascribed Time Slot and Area Code individually to stay away from various quality naming complexity.

Information Transformation: To recognize the patterns and model accuracy, 8-time allotments and various region codes were conceived with regard to quakes susceptibility.

Data Normalisation: The reason for normalization is to change information such that they are either dimensionless and additionally has comparable appropriations. Normalization is a fundamental stage in this study for each data to fit perfectly as each data possess different datatypes. After using scaling normalization, it was formulated in an array for processing.

Figure 2

Procedure illustration



Properties of Earthquake Inventory Map

Earthquake is fairly uncommon for inventories to include data on the date and time of recurrence as well as information on the earthquake's class and morphometry, as well as numerous statistical data and the extent of the earthquake's destruction (Hao et al., 2020; Comert et al., 2019). Earthquake inventory maps are classified into four categories: data on landfall events, earthquake inventory maps, historical data on landfall (Duman et al., 2005), and multi-time earthquake inventory maps (Duman et al., 2005). In addition to field study, earthquake inventories are best created utilizing data acquired from a variety of remote sensing platforms, including space, air, and ground-based platforms (Van Den Eeckhaut et al., 2012). Different procedures are utilized to evaluate the collected data when the data set is considered. Shortly after the occurrence, a research inquiry was carried out. The location, volume, circumstances contributing to the earthquake, kind, and devastation of the earthquake were all meticulously recorded (Comert et al., 2019). Field studies are commonly used to map tiny earthquakes that happened naturally or on a certain day as a result of a specific occurrence.

Hao et al., 2020; Comert et al., 2019) used stereoscopic or manual drawing and visual evaluations to sketch the earthquake regions generated before and after the earthquakes in the region, which were subsequently processed on topographic maps (Van Den Eeckhaut et al., 2012; Duman et al., 2005). There are other ways for interpreting images, such as automatic image recognition (AIR) assessment (Kusak et al., 2019; Hao et al., 2020; and Comert et al., 2019) were utilized to assess the pictures acquired. To build up the earthquake inventory database utilized in this study, these approaches were integrated with other relevant spatial mining algorithms.

Frequency Spectrum Analysis

However, dispersion of frequencies relative to one another, which is connected to a probability distribution, which is frequently used in statistics, depicts the proportion of all observations associated with each value or class of values as a percentage of all observations. (Legorreta et al., 2014). For the slope, as demonstrated in Equation 2, and data aspect, as shown in Equation 3, the generic relative frequency distribution formula was altered and utilized only for this study.

$$F_{Ni} = \frac{f_{Ni}}{\sum f_{Ni}}, \quad (2)$$

Where:

i= {L, NL, VGS, GS, MS, SS, VSS, ES, StS, and VStS}

F_{Ni} : Percentage of the relative frequency of slope type

f_{Ni} : Frequency of slope type

$\sum f_{Ni}$: Total frequency of slope type

$$F_{Nj} = \frac{f_{Nj}}{\sum f_{Nj}} \quad (3)$$

Where:

j = {N, NE, E, SE, S, SW, W, and NW}

F_{Nj} : Percentage of the relative frequency of aspect type

f_{Nj} : Frequency of aspect type

$\sum f_{Nj}$: Total frequency of aspect type

Association Rule Spatial mining

The approach for examining the co-occurrence of events and data relationships of an association rule of various sorts (Can et al., 2009). In terms of probability, association rules indicate the occurrence of occurrences and establish their linkages. The main goal of this approach is to uncover interesting data correlations. This technology is now employed in a variety of industries, including Web use, continuous manufacturing, intrusion detection, and bioinformatics, among many more fields. For constructing rules, there are methods such as Spatial mining (Legorreta et al., 2014). It includes Eclat and FP-growth, SETM, Partition, RARM– Rapid Association Rule Mining, and CHARM, which stands for Charismatic Association Rule Mining.

The Spatial mining processes are as follows:

1. The term "association spatial mining" extracting rules from a data set is referred to as from I = { i_1 , i_2 ,…,i_n } and D = { t_1, t_2 ,…,t_m }, I is a set of n binary characteristics referred to as the database is a collection of transactions known as an item. D comprises a subset of the items in *I,* and each transaction has its own transaction ID.
2. The following is the definition of a rule: 's implication*:* $X \Rightarrow Y$ $X, Y \subseteq I$. Only a set and a single item are specified by a rule, X ⇒ ij for ij ∈ I. Every rule is made up of two different groups of things, referred to as item sets, X and Y…, X is the antecedent, while Y is the consequent.
3. Restrictions imposed on a wide range of options metrics of relevance are employed, as well as the terms and interest in selecting interesting rules from the collection of all viable rules. Degrees of support and confidence at which a person may feel safe and secure is the most well-known limitations.
4. Assume *X*, *Y* to be the item sets, $X \Rightarrow Y$, then T is a set of transactions in a database, and A is an association rule. Three measures of a connection or rule are support, confidence,

and lift (Can et al., 2009). Three frequently used metrics of association may be calculated using these data.

5. The frequency with which the item set appears in the data set is shown by support Equation (4). The percentage of transactions that make up data itemset X is defined as the support of X concerning T.

$$Supp(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|} \quad (4)$$

6. Confidence, as demonstrated in Equation (5), is a measure of how frequently the rule h as been proven correct. The proportion of transactions that include X and also contain Y is the confidence value of a rule, X ⇒ Y, with regard to a collection of transactions, T.

The following are some definitions of confidence:

$$conf(X \Rightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)} \quad (5)$$

The lift as indicated in Equation (6) of a rule, is defined as:

$$Lift(X \Rightarrow Y) = \frac{Supp(X \cap Y)}{Supp(X) * Supp(Y)}, \quad (6)$$

assuming X and Y were independent, the ratio of observed to anticipated support. Adapting the Spatial mining method (Alptekin et al., 2019) to the spatial data mining sector was applied for the earthquake area in this work. When looking through the literature, it was discovered that the Spatial mining method had been employed in earthquake investigations (Kusak et al., 2019). Using WEKA 3.8 version software, the as part of the research area's hydrology and NDVI category data was analyzed using the Spatial mining method.

Machine Learning Approach Algorithms based on SVM and K-means

This subsection depicts the ML techniques that will be used to predict the quake susceptibility. Below more discussion will be given on the utilized ML techniques.

K-means: The technique of splitting data into groupings is known as cluster analysis. The classifications are not preset, unlike the categorization approach. The similarity of the given data determines which data will be separated into several categories and clusters for analysis, as well as they will be separated into how many different groups. The clustered analysis is used in a wide variety of fields such as biology and medicine as well as anthropology and marketing.

The K-means include partitioning, and the most effective clustering method is used, known as the square-error criterion algorithm. The overall goal is to find the division that minimizes the total square error for a given number of clusters. The following are the K-means procedures used in this study:

1. In n-dimensional space, the data collection contains N samples.
2. One of the approaches, such as Elbow or Silhouette, was used to calculate the number of clusters (K). C_1, C_2, C_3.., C_k are the K clusters of N samples each.
3. The members' allocation to the K centers was already determined (Hao et al., 2020; Comert et al., 2019). There are n_k samples per C_k. of which belong to precisely one cluster, therefore, $\sum n_k = N$ where: k=1, 2,…, K.
4. In cluster C_k indicative in Equation (7), the mean vector M_k is defined as the cluster's centroid or where x_{ik} cluster Ck's twelfth specimen is it?

   $$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} x_{ik} \quad (7)$$

5. The total of the squared differences in C_k between each sample and its centroid (Hao et al., 2020; Comert et al., 2019) in Equation (8), known as the within-cluster variation (ek2), the square error (Hao et al., 2020; Comert et al., 2019).

   $$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2, \quad (8)$$

6. Equation (9), which shows the total of the within-cluster variances, is known as the within clusters sum of square error and represents the square error over the whole clustering space comprising K clusters. (E_k^2).

   $$E_k^2 = \sum_{k=1}^{K} e_k^2, \quad (9)$$

7. Using a square-error clustering approach, the goal is to discover a partition with K clusters that minimizes E_k^2 for a given K (Kusak et al., 2019). The Euclidean distances formula in Equation (10) is used to calculate the elements of each C_k set as well as their distances from the M_k center.

$$D(M_k, X_i) = D_{mk} = \sqrt{\sum_{i=1}^{K} (M_k - X_{ik})^2}, \quad (10)$$

8. For instance, the sample (x_{ik}) is regarded as an element of that cluster if it is near one (M_k or M_{k+1}) of the two cluster centers. Then, $D(M_k, X_i) < D(M_{k+1}, X_i)$ *is* $X_i \in C_k$.
9. All cluster elements have their Euclidean distances computed, and their cluster membership is established.
10. Cluster members and centers may be separated by a variety of distances, the transactions are updated repeatedly until the centers do not change (Kusak et al., 2019).

Categorical data analysis does not use K-means, although the method has been made public in a variety of investigations (Kusak et al., 2019). The K-means method may also be utilized for categorical data using the WEKA version 3.8 software systems (Comert et al., 2019). In geographical data mining, clustering is utilized, whereas, in earthquake research, the K-means method is used (Hao et al., 2020; Comert et al., 2019). An analysis of elevation, slope, aspect, stream, and NDVI data was carried out using WEKA 3.8 software and the K-means technique. The data was represented graphically using ArcGIS 10.5 software.

Support Vector Machine (SVM): SVM is effective, especially in high dimensional, since they use a subset of focuses on support vectors (Tuncal et al., 2020). It is known as a supervised learning algorithm; it is used for characterization and relapse arrangements. It involves theoretical and numeric abilities to deal with the relapse issue. It gives the most imperative accuracy rate while doing an estimate of the immense dataset. It is a solid ML strategy that depends on 3D and 2D demonstration. SVM groups basic features from all categories; these features are called support vectors. SVM (Support Vector Machine) is a directed grouping model generally applied in regression and classification areas (Sekeroglu and Tuncal, 2020; Sekeroglu and Emirzade, 2018). The Linear classifier of SVM intends to augment the separation between the hyperplane and the closest information point, known as the restricted distance. SVM has a great precision of up to 98% on information analysis and 78.35% precision utilizing the polynomial portion.

The SVM depends on the idea of choice planes that characterize decision limits. A classification process typically includes preparing and testing information which comprises some instances. SVM is a helpful procedure for information classification. Each occurrence in the preparation set contains one "target label" (class marks) and a few "attributes" (highlights).

Given a preparation set (x_i, y_i), i=1,…,l where xi ∈ Run and y ∈ (1,- 1)l, the SVM needs the solution for optimization in Equation (11) and Equation (12)

$$\text{Min}_{w,\, b,\, £}\ 1/2w^T\ w + c\ \Sigma^1{}_{i=1\varepsilon I} \qquad (11)$$

$$\text{Subject to } y_i\ (w^T\ Ø(x_i) + b) > 1_{-\varepsilon I},\ \varepsilon_i >= 0. \qquad (12)$$

Afterwards, SVM looks for a linear approach isolating hyper plane with the maximal space in a higher-dimensional space. Here, the vectors xi is processed into a higher dimensional space by the variable Ø. Moreover, k (x_i, x_j) = Ø (x_i) Ø (x_j) is known as the kernel variables.

These involve a linear polynomial, RBF, and sigmoid. Several parts can be utilized in SVM models given in Equation (13)

$$Ø= \{x_i{*}x_j \text{ linear } (\gamma x_i x_j+\text{coeff})^d \text{ polynomial Exp } (-\gamma|x_i-x_j|^2)$$

$$\text{RBF= Tanh } (\gamma x_i x_j+\text{coeff}) \text{ sigmoid}\} \qquad (13)$$

There is a similar connection between the Radial Basis Function (RBF) and SVMs classifiers. The RBF is by far the most mainstream decision of part types utilized in SVM. In the area of this thesis imaging, the important utilization of SVMs is in rain discovery. The main SVM is a linear classifier. Similarly, nonlinear SVMs can be made. The component space is a nonlinear guide from the main information space. They are developed by finding a bunch of planes that differentiate at least two classes of information. By developing these planes, SVM finds the limits between the information classes; the components of the information that characterize these limits are called support vectors.

Radial Basis Function is given in Equation (14):

$$K\ (x, x') = \exp\ (-|x-x'|^2 / (2\sigma^2\ )). \qquad (14)$$

The altered kernel is utilized to get the last classifier. The piece is then adjusted in the information path by utilizing the gotten support vectors.

Metrics of Performance

The clustering analysis was validated using the performance metrics technique. The sample size method equation was used to determine in the equation, the number of samples required to represent each cluster given in Equation (15),

$$n = \frac{nt^2p(1-p)}{d^2(N-1)+t^2p(1-P}, \tag{15}$$

where:

n: Sample size

N: Population size

t: Standard value of 1.96 in confidence level at 95%

p: population proportion (0.5)

d: the margin of error

Equation (16) is used to evaluate the performance metrics for accuracy; then, the recall is evaluated using Equation (17), precision using Equation (18) as well as F1 score Equation (19) (Sekeroglu, 2004).

$$Accuracy = \frac{TP+T}{TP+TN+FP+F}, \tag{16}$$

$$Recall = \frac{TP}{TP+FN}, \tag{17}$$

$$Precision = \frac{TP}{TP+F}, \tag{18}$$

$$F1 - score = 2 * \frac{Pr\,ecision-Re\,ca}{Pr\,ecision+Re\,c}, \tag{19}$$

where:

TP --- True Positive (Earthquakes correctly observed from the 90s to the 20s).

TN --- True Negative (Non-earthquake correctly observed from the 90s to the 20s).

FP --- False Positive (Earthquake correctly observed from the 90s to the 20s).

FN --- False Negative (Non-earthquake correctly observed from the 90s to the 20s).

For metric performance analysis, random points from each cluster were chosen at random. These points' metric values were calculated using the dataset discussed in the previous section, which includes data on latitude and longitude.

CHAPTER IV

Result and Discussion

The results shown in this resulting section show the result of the experiment after performing the spatial mining techniques and various other statistical approaches to validate the performance of the models.

The generated maps of earthquake susceptibility and its spatial performance

We proceed with a qualitative assessment and comparison with various requirements. A configuration was made to the data that returned a relevant number of countries with consequent earthquakes, and these countries are qualitatively evaluated, considering the spatial data mining technique, grouping them based on the earthquake occurrence. More emphasis will be made in the accompanying paragraphs based on these quakes and the country of occurrence. Figure 3 depicts the heat map of the dataset as a whole. This data shows the quake's occurrence around the world; here, spatial mining will be applied to this data. However, the clustering approach was also applied to show each region adequately. Table 4 portrays a section of the input quake data. The process utilized a comparable area of enlightening assortments.

The quakes susceptibility maps were obtained by clustering, as shown in Figure 3 discussed above shows different spatial distributions on the various continent on the maps. On the map, the Asian continent projected recurrent quakes, showing an overestimation of quakes susceptibility and a lack of ability to separate prone areas from stable areas. Every component of spatial mining accepts a key part to getting a higher analysis of the quake occurrence. To survey the sufficiency of our strategy, we tried the dataset on different parameters for a clearer view of the spatial models. The correlation between each region of the dataset will be discussed further in the accompanying sub-sections. This concept shows each dataset's similarities. It shows that the relationship between the mean range and the edge is solid, the mean border and region, and the connection between each quake's highlights.

Figure 3

Depicting the earthquake data clustering

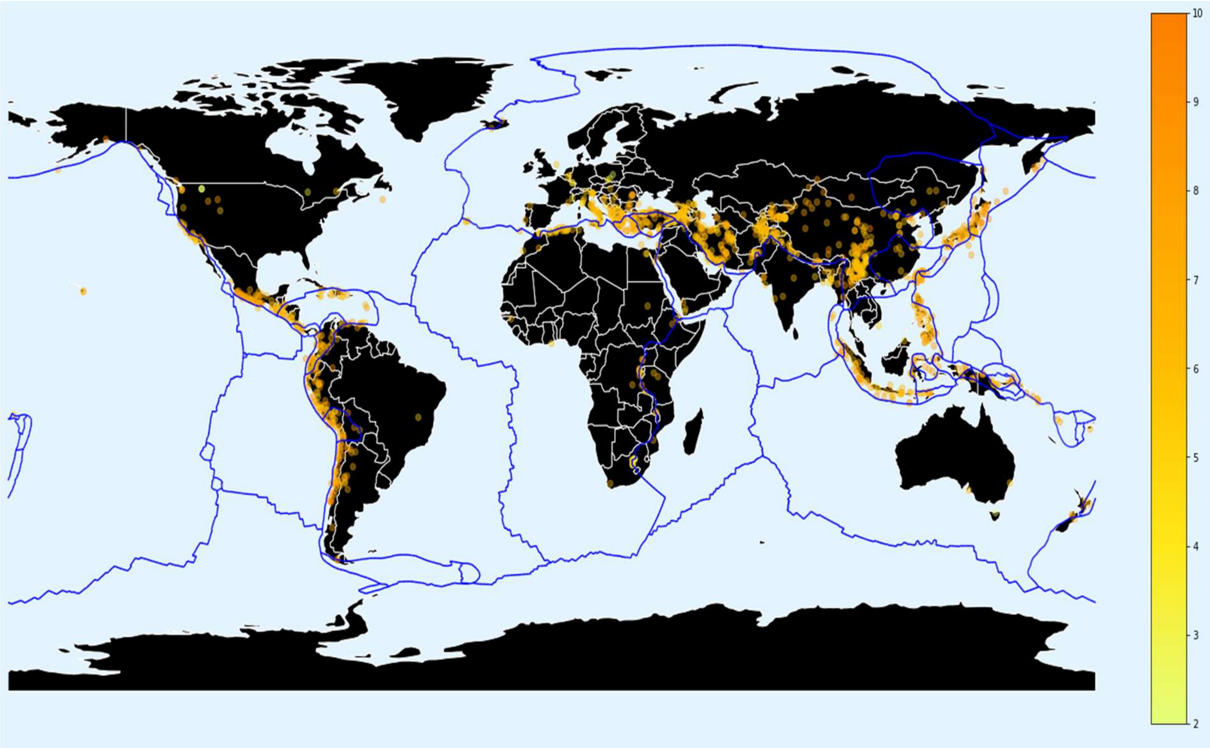


Table 4.

Quake data depiction

latitude	longitude	depth	mag	magType	nst	gap	dmin	rms
37.0855	-98.040667	6.36	1.56	ml	12	137	0.05775	0.04
62.7988	-149.4673	6	0.9	ml	NaN	NaN	NaN	0.63
32.966833	-115.56283	9.88	2.56	ml	47	75	0.06521	0.24
38.4284	-118.8892	5.5	1.2	ml	6	210.16	0.072	0.0885
38.4284	-118.9013	6.7	1.9	ml	13	70.02	0.072	0.1497
37.0855	-98.040667	6.36	1.56	ml	12	137	0.05775	0.04
62.7988	-149.4673	6	0.9	ml	NaN	NaN	NaN	0.63
32.966833	-115.56283	9.88	2.56	ml	47	75	0.06521	0.24
38.4284	-118.8892	5.5	1.2	ml	6	210.16	0.072	0.0885

Validation of earthquake susceptibility mining in the Study Area

Regional mining was utilized to assess the viability of several countries' vulnerability to quakes occurrence made by the spatial approach, as depicted in Figure 4 As discussed in the above section, Figure 5 shows the magnitude of the quake per year rate. Vulnerability appraisals were basically scattered between several country limits, as shown in the figure discussed above. The spatial dissemination of the country's death rate was significantly higher in Indonesia than in different countries. The spatial mining assisted in differentiating each country concerning the accuracy, review, responsiveness, particularity, and all-out exactness. Countries with death size greater than 10000 deaths have unfortunate order plans for quake susceptibility. Contingent upon the inspecting size used and the region picked, a few years with quakes occurrence maps were made as discussed in the image above. The examining size and region in this examination brought about different commitments to the mining process. In 1976 and 2004, the inspecting size and region picked may have affected different expectation capacities. The depicted proportion shows a ratio of 6:3 in the occurrence of death and focuses on a proper way to expand the immediate planning forecast for reoccurrence. Spatial mining strategies could be utilized to explore the portrayal expected for making crude information expectations, as shown in the discussion above. Utilizing dimensionality decreased the mining effectiveness determination of spatial examples and diminished the number of regional boundaries.

Subsequently, Figure 6 shows the frequency of the quake magnitude. One expected line of exploration is to investigate the chance of utilizing a strong approach to determine the frequency to analyze each occurrence for future reoccurrence. Furthermore, different sample procedures in quake susceptibility mapping have differing degrees of accuracy. Therefore, sampling strategies should be considered when using spatial mining techniques. The depiction of the statistical value analysis is given in Table 4.2. It shows the standard deviation and means of the resulting quakes. Table 5 shows the statistical analysis of the quake resulting data. This factual examination shows logical intrusion that depicts, gathers, and investigates the information of the quakes as well as the patterns to recognize any significant data change in the quakes. The table depicts the mean, median, and standard deviation of the rainfall, which can be utilized as unmistakable measurable strategies for the spatial model to change crude perceptions into what can be shared and comprehended.

Figure 4

Depicting the earthquake countries' magnitude vs. death rate

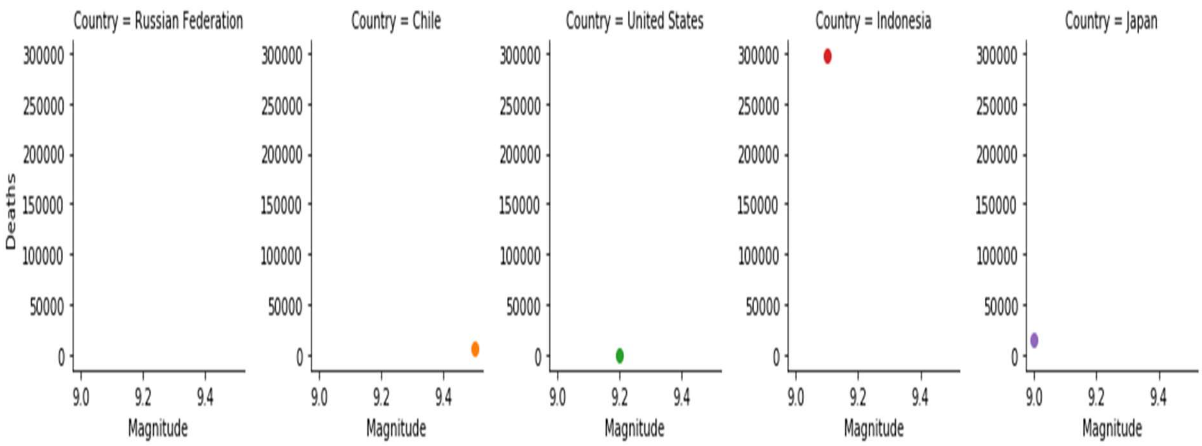


Figure 5

Depicting the earthquake year magnitude vs. death rate

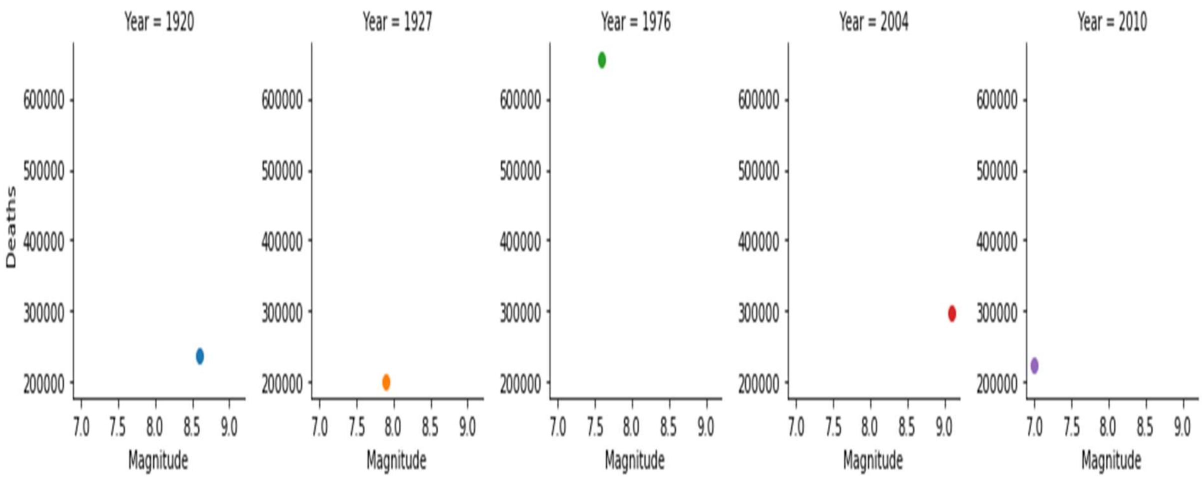


Figure 6

Depicting the earthquake frequency magnitude

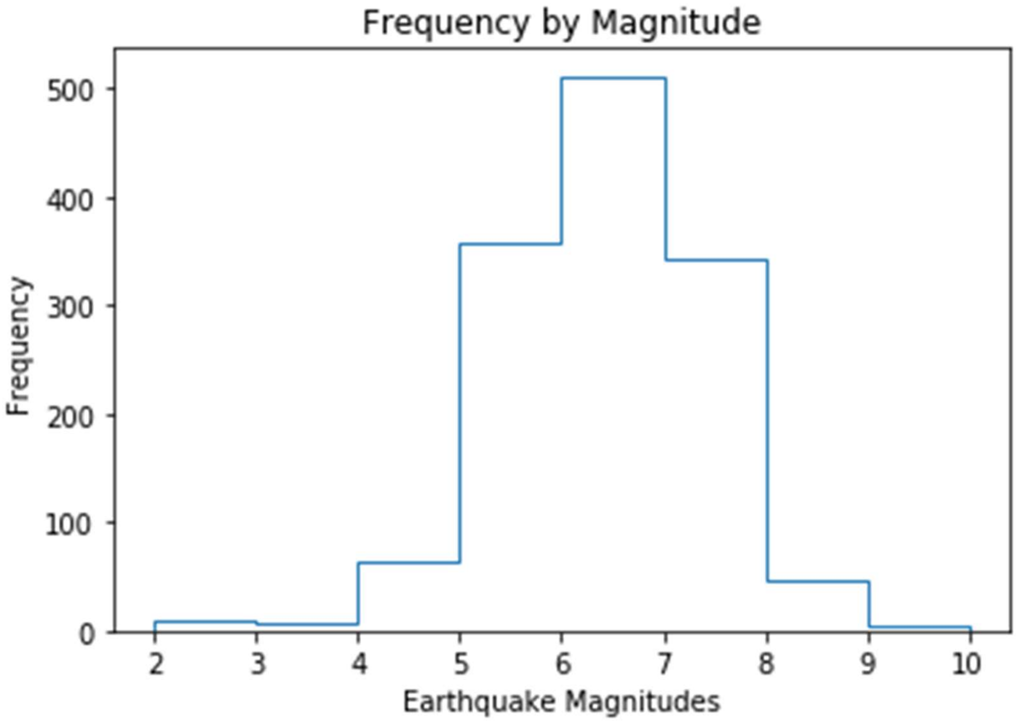


Table 5.

Statistical report of the Quakes

	Depth	Lat	ML
count	9639.000000	10639.000000	10639.000000
mean	6.500000	49.654606	2.50000
std	6.456055	63.397233	1.11933
Min	0.000000	0.000000	1.00000
25%	2.750000	0.000000	1.75000
50%	5.500000	22.000000	2.50000
75%	9.250000	83.850000	3.25000
max	42.000000	400.000000	4.00000

Earthquake mining factors rate in the Study Area

The quake influencing components in a given exploration district is depicted and analyzed, and it's difficult to express which of the geographical, land, hydrological, and distance-to-factors are the most huge and essential in respect to the magnitude. This can be taken care of by introducing the spatial mining technique, as depicted in Figure 7 This depicts its logarithmic value frequency in order to predict and analyze each magnitude to be expected for reoccurrence. Quakes might repeat under conditions like past reoccurrence, as per the earth’s vulnerability, or vice versa. Along these lines, for instance, size 6.5 quakes happen multiple times more oftentimes than magnitude 7's and multiple times more frequently than extent 8's. This example is known as the law of power distribution. Here would seem for each increment of one point in greatness, the quake becomes multiple times less incessant. For instance, assuming we realize that there were 15 quakes somewhere in the range of 5.0 and 5.9 in a specific locale in a time of 70 years, that works to around one quake in three years. We can utilize this to moderately work out the likelihood that a quake will hit a specific area, in spite of the fact that it is difficult to know precisely when. Following this dispersion above, we can "anticipate" that a tremor estimating somewhere in the range of 6.0 and 6.9 ought to happen about once at regular intervals around here.

Input factors for quakes mining were picked from several countries, as depicted in Figure 8 It shows the rate of death in the selected countries. Each country analyzed in the figure shows that Japan had the highest death rate based on the figure. There is a need to indulge in preventive measures to prevent the death toll in this locale. Geological boundaries, for example, elevation, slant point, plan curve, profile bend, and distance to stream, were acquired and assessed utilizing advanced spatial mining and are totally included in performing the experimentation in the figure depicted. The geospatial data authority gave the distance of each locale in the region. The Geological Survey gave topographical components the view that the Asian continent is susceptible to the reoccurrence of quakes with high magnitudes.

Figure 7

Depicting the earthquake frequency scale of the world

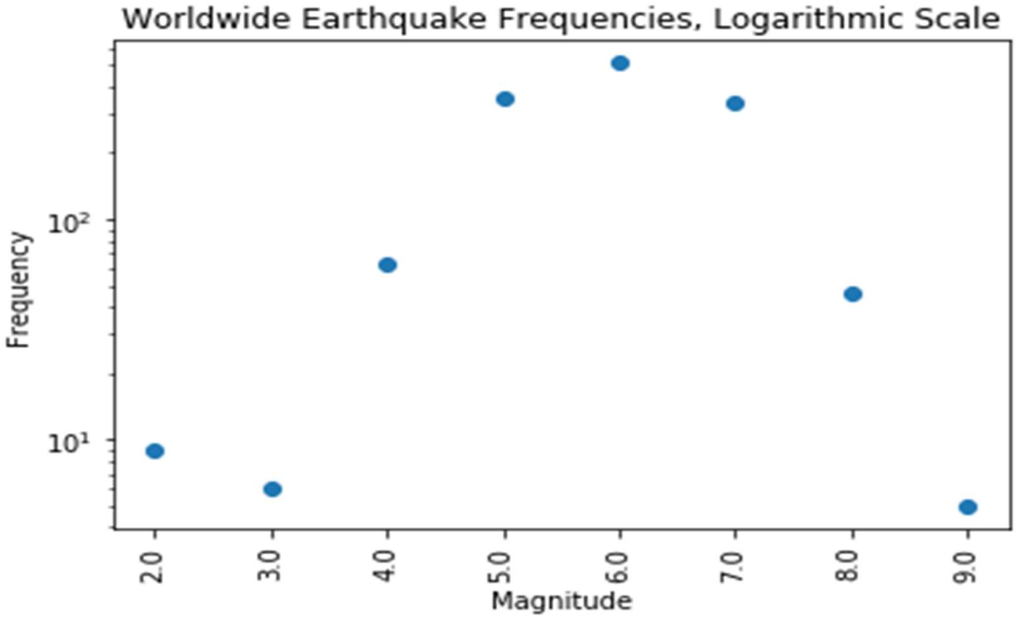


Figure 8

Depicting the earthquake yearly death rate and magnitude

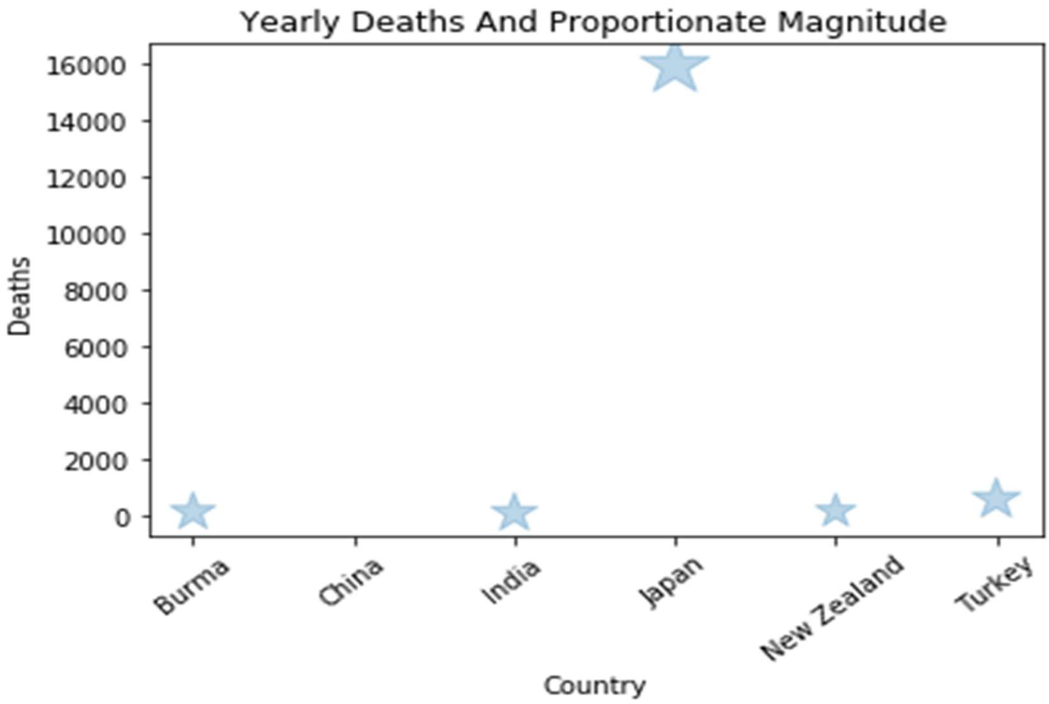


Figure 9 shows the quakes magnitude and its death toll. This figure shows that the magnitude might sometimes instigate the death toll. It shows that quakes with a magnitude of 8 had a higher death toll compared to the quake of 9 magnitudes. These results give rise to the question that areas with magnitude nine quakes were more prepared in the case of any reoccurrence. It is displayed in the recurrence proportion of all subclasses of every quake influencing part of the world. The mining result shown in the figure allows every country with quake reoccurrence with higher magnitude may affect the death toll significantly in the event of an earthquake. Figure 10 shows further analysis of the magnitude per year for three known countries with major quakes. To diminish model clamor for quake occurrence, the exploration has exhorted those connections between affecting countries to show the reach of its magnitude, and preventive measures can be considered. The countries can endure some collinearity peculiarities among avalanche influencing components. It shows that the US had a higher magnitude in the 60s. These quake influencing countries give valuable information for making quake maps, which depend on a recurrence proportion-based quantitative estimation.

Figure 9

Depicting the earthquake magnitude per death rate

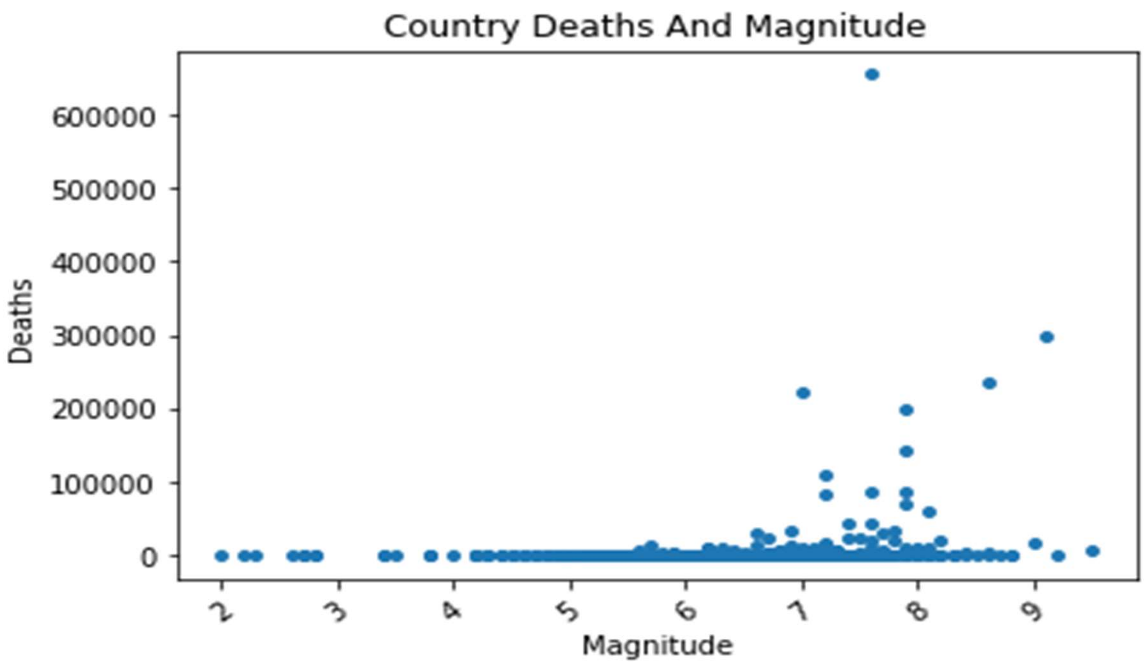
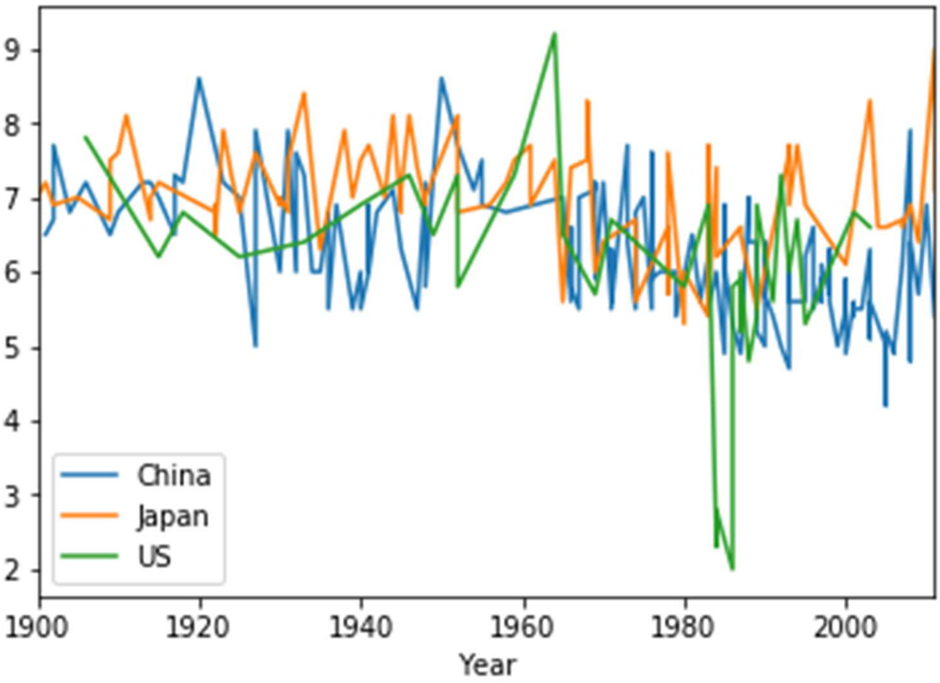


Figure 10

Depicting the earthquake magnitude in the three main countries with quake occurrence



Discussion of Findings and Sampling of ML Model

To guarantee the consistency of each used technique, the experiment was executed similarly. We have utilized spatial data mining were applied to a relatively immense informational index of earthquakes around the world. The used approach in this hypothesis provided a promising presentation as depicted in this area. The made request structure is shown to be good for making quake analysis capable. Similarly, due to the differentiation in preprocessing, the outcome was viable for every experimentation carried out. Figure 11 shows six major countries with the highest magnitudes of quakes. Each country was linearly separable. The accuracy relationship in the correlation with the earthquake magnitude is given in Figure 12 for concluding the quake analysis. From Figure 12, performing effectively, the correlation analysis shows the world has seen more quakes around the magnitude range of 6 and 7. Consequently, the result derived from this test shows that experts can use spatial mining proficiently in anticipating and analyzing quakes around the world effectively at whatever point used. Note that the distinction in precision was somewhat similar for the countries discussed in the figure above. This is a reminder to all researchers and practitioners to use the proposed strategy.

To guarantee consistency of the spatial data mining technique, similar ML clustering techniques were tried in order to deduce how accurate the result would be. We have utilized two models, K-means and SVM. They are known for their classification and clustering power, as discussed in the literature section. An amount of 10000 data was used for this experiment. The data was split into train and test sets to show the presence of precision in the quake data for analysis and forecast. The data were all executed in this experiment using a train/test parting of 80% to 20% extent. Several parameter tuning was implemented, and the best parameters were selected. After optimizing it with various folds of cross-validation as discussed in the methodology, we were able to achieve a better result. The accuracy relationship for the ML model for concluding the quake accuracy analysis prediction is given in Table 6. It shows the susceptibility of the quake data, classifying if the directed location will be susceptible or not susceptible. Of the models performing effectively, the best model with the highest accuracy is the SVM getting a precision of 96%, while the K-means settled at 93%. This result can be to the conclusion that spatial data mining plays a predominant role in earthquake analysis and diminishing disasters profoundly.

Figure 11

Depicting the earthquake magnitude for different countries

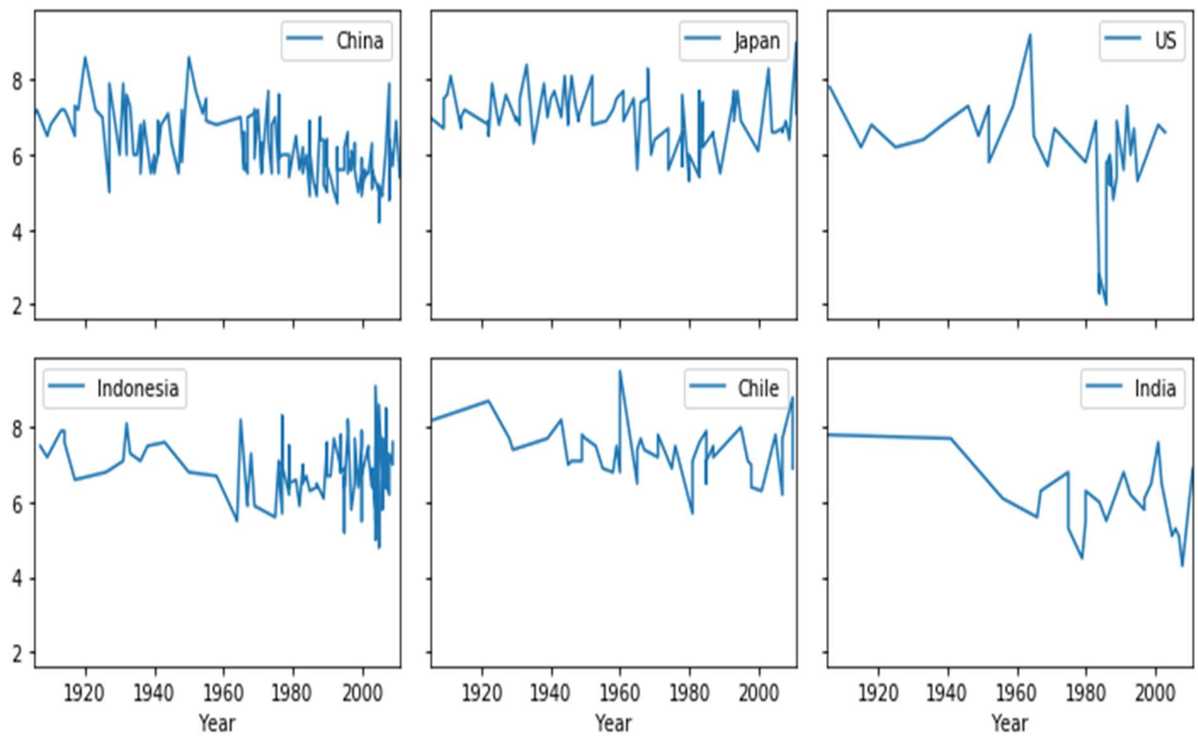


Figure 12

Depicting the earthquake magnitude correlation

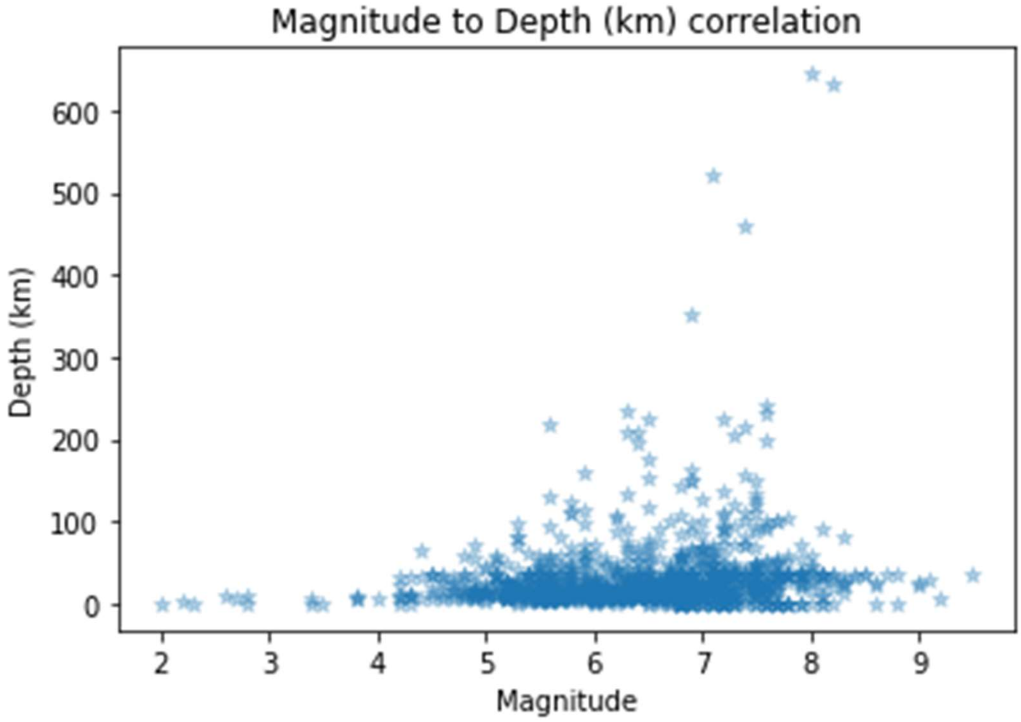


Table 6.

Experimental result of the used model with others (%)

Authors	Recall	Precision	F1 score	Accuracy
Supriyadi et al. (2018)	-	-	-	0.79
Asim et al. (2020)	0.81	0.91	0.89	0.82
Yu, X. (2017)	0.83	0.89	0.87	0.89
This thesis (SVM)	0.85	0.98	0.92	0.96
This thesis (K means)	0.87	0.95	0.91	0.93

CHAPTER V

Conclusion and Recommendation

Conclusion

The review region was picked as a contextual analysis for earthquakes created by quakes involving around the world vulnerability planning. The outcomes were approved utilizing objective models like applying spatial data mining, recall, and precision based on a clustering approach to pinpoint the exact locations. The discoveries of the analysis lead to the accompanying conclusions.

- The precision and recall predictions derived using the suggested spatial mining for ML model performed well in terms of precision and accuracy.
- Different sample procedures were enhanced for susceptibility of quakes assessing the accuracy, which should be studied and compared to recall and precision on geodata.
- For the different models, inspecting the proportion of 1:2 of earthquake and non-earthquake created in the mining process delivered the best outcomes utilizing the geo-mapping, particularly where earthquakes were set off by the tremor.

Finally, the experimental results show that using the spatial mining learning model with ML models improves the accuracy of the prediction of quake mapping susceptibility. The earthquake susceptibility accuracy of the maps in this study could be improved in the future by choosing the best sampling approach and looking at highly efficient deep learning techniques. Of the models performing effectively, the best model with the highest accuracy is the SVM getting a precision of 96%, while the K-means settled at 93%. This result can be to the conclusion that spatial data mining plays a predominant role in earthquake analysis and diminishing disasters profoundly. Organizers, engineers, and decision-makers in catastrophe arrangements might observe the quakes mapping created in this thesis important in decreasing monetary casualties and losses.

Recommendations

Our existing model might be improved in a number of ways. However, the study suggests that training a rotation-invariant model with data augmentation and comparing its performance to that of a locally aligned convolutional neural network may be interesting. The study can optimize over various distances to evaluate which ones are best for extracting features, or it can even formulate the problem using soft attention to align the filters rather than hard attention on a predefined set of distances. Other possible extensions include finding better ways to balance the data, proposing better metrics for comparing different susceptibility maps, using different models for landslide susceptibility mapping such as auto-encoders, and treating the ground truth as positive and unlabeled data rather than positive and unlabeled data. Another intriguing option suggests that this work be expanded to include learning the uphill direction without having it as a significant prior in the model.

References

Abidine, R. Z., & Abdelmansour, N. (2019). Landslide susceptibility mapping using information value and frequency ratio for the Arzew sector (North-Western of Algeria). *Maden Tetkik ve Arama Dergisi*, (159), 1-10.

Adeli, H., & Panakkat, A. A (2009). probabilistic neural network for earthquake magnitude prediction. *Neural Netw*., 22, 1018–1024. [CrossRef]

Aditian A, Kubota T, & Shinohara Y (2018). Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* 318:101-111.

Amezquita-Sanchez, J., Valtierra-Rodriguez, M., & Adeli, H. (2017). Current efforts for prediction and assessment of natural disasters: Earthquakes, tsunamis, volcanic eruptions, hurricanes, tornados, and floods. *Sci. Iran*., 24, 2645–2664. [CrossRef]

Amit, S.N.K.B., & Aoki, Y. (2017). Disaster detection from aerial imagery with convolutional neural network. *In Proceedings of the 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Surabaya, Indonesia*, 26–27 September, pp. 239–245.

Ashiquzzaman, A., Oh, S.M., Lee, D., Lee, J., & Kim, J. (2021). Context-aware deep convolutional neural network application for fire and smoke detection in virtual environment for surveillance video analysis. *In Smart Trends in Computing and Communications, Proceedings of the SmartCom 2020, Paris, France*, 29–31 December 2020; Springer: Berlin/Heidelberg, Germany, pp. 459–467.

Asim, K. M. Idris, A. Iqbal T. & Martínez-Álvarez, F. (2018). Earthquake prediction model using support vector regressor and hybrid neural networks, *PLoS One, vol. 13, no. 7,* *pp. 1-22.*

Barlow J, Martin Y, & Franlin SE. (2003). Detecting trans lational lands lide scars using segmentation of Landsat ETM+and DEM data in the northern Cas-cade Mountains, British Columbia. *Canadian journ al of remote sensing*; (29), p. 510-517.

Boonsuk, R., Sudprasert, C., & Supratid, S. (2019). An Investigation on Facial Emotional Expression Recognition Based on Linear-Decision-Boundaries Classifiers Using Convolutional Neural Network for Feature Extraction. *In Proceedings of the 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), Pattaya, Thailand*, 10–11; pp. 1–5.

Brooks Nick, & Adger W Neil (2013). Country level risk measures of climate-related natural disasters and implications for adaptation *to climate change. Clim Res. Tyndall Centre for Climate Change*;. p. 1–30.

Bui, D. T., Lofman, O., Revhaug, I., & Dick, O. (2011). Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Natural hazards*, 59(3), 1413.

Cavallo, A. U. Troisi, J. & Forcina M. (2021). Texture analysis in the evaluation of Covid-19 pneumonia in chest X-ray images: a proof of concept study," *Current Medical Imaging*, vol. 17, no. 9.

Chalkias, C., Ferentinou, M., & Polykretis, C. (2014). GIS-Based Landslide Susceptibility Mapping. *Geosciences*.

Chen, Y., Zhang, Y., Xin, J., Wang, G., Mu, L., Yi, Y., Liu, H., & Liu, D. (2019). UAV Image-based Forest Fire Detection Approach Using Convolutional Neural Network. *In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, 18–21; pp. 2118–2123.

Datar Ashlesha, Liu Jenny, Linnemayr Sebastian, & Stecher Chad (2013). The impact of natural disasters on child health and investments in rural India. *Soc Sci Med* 2013;76:83–91.

Direkoglu, C. (2020). Abnormal Crowd Behavior Detection Using Motion Information Images and Convolutional Neural Networks. *IEEE Access*, 8, 80408–80416. [CrossRef]

Dong, L., (2009). Review of Landslide Geological Disaster Monitoring Method. *Mining Technology. 2009,9(3),63-65.*

Echtioui, A. Zouch, W. Ghorbel, M. Mhiri, C. & Hamam, H. (2020) Detection methods of COVID-19," *SLAS Technology*, vol. 25, no. 6, pp. 566–572.

Gonzalez, A., Zuniga, M.D., Nikulin, C., Carvajal, G., Cardenas, D.G., Pedraza, M.A., Fernández, C., Munoz, R., Castro, N., & Rosales, B.. (2017). Accurate fire detection through fully convolutional network. *In Proceedings of the 7th Latin American Conference on Networked and Electronic Media (LACNEM 2017)*, Valparaiso, Chile, 6–7 November.

Hartawan, D.R., Purboyo, T.W., & Setianingsih, C. (2019). Disaster Victims Detection System Using Convolutional Neural Network (CNN) Method. *In Proceedings of the 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia*, 1–3, pp. 105–111.

Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A., & Linh, N.T.T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci. Front.*, 12, 101075. [CrossRef]

Khanh, N. Q. (2009). Landslide hazard assessment in muonglay, Vietnam applying GIS and remote sensing. *Dr. rer. Nat. at the Faculty of Mathematics and Natural Sciences Ernst-Moritz-Arndt-University Greifswald.*

Kradolfer, U. (2013). SalanderMaps: A rapid overview about felt earthquakes through data mining of web-accesses. *In Proceedings of the EGU General Assembly Conference, Vienna, Austria*, 7–12 April; pp. EGU2013–6400.

Layek, A.K., Poddar, S., & Mandal, S. (2019). Detection of Flood Images Posted on Online Social Media for Disaster Response. *In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, Gangtok, Sikkim, India, 28–28 February 2019; pp. 1–6.

Lee, W., Kim, S., Lee, Y.-T., Lee, H.-W., & Choi, M. (2017). Deep neural networks for wild fire detection with unmanned aerial vehicle. *In Proceedings of the 2017 IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 8–11; 252–253.

Li, T., Zhao, E., Zhang, J., & Hu, C. (2019). Detection of Wildfire Smoke Images Based on a Densely Dilated Convolutional Network. *Electronics*, 8, 1131. [CrossRef]

Madichetty, S., & Sridevi, M. (2019). Detecting informative tweets during disaster using deep neural networks. *In Proceedings of the 2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, Bangalore, India, 7–11; pp. 709–713.

Mangalathu, S., & Burton, H.V. (2019). Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions. *Int. J. Disaster Risk Reduct.*, 36, 101111. [CrossRef]

Mantovani F, Soeters R, & Western CJ. (1996). Remote Sensing Techniques for Landslide Studies and Disaster Zonation in Europe. *Geomorphology;*, 15(3-4), p. 213-225.

Martin YE, & Franklin SE. (2005). Classification of soil-and bed rock-dominated landlsides in British Colunbia using Martin YE, Franklin SE. Classification of soil-and bed rock-dominated landlsides in British Colunbia using segmentation of satellite imagery and

DEM data. *International journal of remote sensing;*, 26(7), p. 1505-1509.

Merz, B., Kreibich, H., & Lall, U. (2013). Multi-variate flood damage assessment: A tree-based data-mining approach. *Nat. Hazards Earth Syst. Sci*., 13, 53–64. [CrossRef]

Mignan, A., & Broccardo, M. (2020). Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. *Seism. Res. Lett.*, 91, 2330–2342. [CrossRef]

Nguyen, D.T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage assessment from social media imagery data during disasters. *In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Sydney, NSW, Australia, 31 July–3 , pp. 569–576.

Nimmagadda Shastri L, Dreher Heinz (2007). Ontology based data warehouse modeling and mining of earthquake data: prediction analysis along Eurasian–Australian continental plates. *In: 5th IEEE international conference on industrial informatics*; 2007. p. 597–602.

Nohani, E., Moharrami, M., Sharafi, S., Khosravi, K., Pradhan, B., Lee, S., & M. Melesse, A. (2019). Landslide susceptibility mapping using different GIS-based bivariate models. *Water,* 11(7), 1402.

Nunavath, V. & Goodwin, M. (2018). The role of artificial intelligence in social media big data analytics for disaster management-initial results of a systematic literature review. *In Proceedings of the 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Sendai, Japan, 4–December; pp. 1–4.

Padmawar, P.M., Shinde, A.S., Sayyed, T.Z., Shinde, S.K., & Moholkar, K. (2019). Disaster Prediction System using Convolution Neural Network. *In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 17–19; pp. 808–812.

Pourghasemi, H. R., Pradhan, B., & Gokceoglu, C. (2012). Remote sensing data derived parameters and its use in landslide susceptibility assessment using Shannon's entropy and GIS. *In Applied Mechanics and Materials* (Vol. 225, pp. 486-491). Trans Tech Publications.

Sahay, R.R. & Srivastava, A. (2014). Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water Resour. Manag*., 28, 301–317. [CrossRef]

Samudre, P., Shende, P., & Jaiswal, V. (2019). Optimizing Performance of Convolutional Neural Network Using Computing Technique. *In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Pune, India, 29–31 March 2019;1–4.

Schlemper, J., Caballero, J., Hajnal, V., Price, A.N., & Rueckert, D. (2017). A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging*, 37, 491–503. [CrossRef] [PubMed]

Sekeroglu, B. (2004). Classification of Sonar Images Using Back Propagation Neural Network, *IEEE IGARSS*, 5, 3092-3095.

Sekeroglu, B. & Emirzade, E. (2018). A Computer-Aided Diagnosis System for Lung Cancer Detection Using Support Vector Machine, *International Workshop on Pattern Recognition*, China, May, , 108282.

Sekeroglu, B., & Tuncal, K. (2021). Prediction of cancer incidence rates for the European continent using machine learning models, *Health Informatics Journal*, https://doi.org/10.1177/1460458220983878, 2021.

Sulistijono, I.A., Imansyah, T., Muhajir, M., Sutoyo, E., Anwar, M.K., Satriyanto, E., Basuki, A., & Risnumawan, A. (2018). Implementation of Victims Detection Framework on Post Disaster Scenario. *In Proceedings of the 2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*, Bali, Indonesia, 29–30 October; pp. 253–259.

Supriyadi, B., Windarto, A. P. & Soemartono T. (2018). Classification of Natural Disaster Prone Areas in Indonesia using K-Means, *Int. J. Grid Distrib. Comput*., vol. 11, no. 8, pp. 87-98.

Tang, C., Zhu, Q., Wu, W., Huang, W., Hong, C., & Niu, X. (2020). PLANET: Improved convolutional neural networks with image enhancement for image classification. *Math. Probl. Eng*.,. [CrossRef

Tonini, M., D'Andrea, M., Biondi, G., Degli Esposti, S., Trucchia, A., & Fiorucci, P. (2020). A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy. *Geosciences*, 10, 105. [CrossRef]

Tuncal, K, Sekeroglu, B., & Ozkan, C. (2020). Lung Cancer Incidence Prediction Using Machine Learning Algorithms, *Journal of Advances in Information Technology*, 11(2), 91-96.

Venkatesan, M., Thangavelu, A., & Prabhavathy, P. (2012). An improved Bayesian classification data mining method for early warning landslide susceptibility model using GIS. *In Proceedings of the Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, Springer, Gwalior, India, 12–16; pp. 277–288.

Wang Zhihua. (1999). Reviewing and Prospecting for Applying Remote Sensing to Landslide and Debrisflow Investigation . *Remote Sensing for Land & Resources*; 1999, 41(3), p.

10-15.

Yang, Liu, Fu Zhengye, Zheng Fengbin. (2015). Review on High Resolution Remote Sensing Image Classification and Recognition. *Journal of Geo-information Science. 2015(9),1080-1091.*

Yang, S., Hu, J., Zhang, H., & Liu, G. (2021). Simultaneous Earthquake Detection on Multiple Stations via a Convolutional Neural Network. Seism. *Res. Lett*., 92, 246–260. [CrossRef]

Yu, X. (2017). Disaster prediction model based on support vector machine for regression and improved differential evolution, *Nat. Hazards*, vol. 85, no. 2, pp. 959-976.

Yuan, F., Zhang, L., Xia, X., Huang, Q., & Li, X. (2019). A wave-shaped deep neural network for smoke density estimation. IEEE Trans. *Image Process*., 29, 2301–2313. [CrossRef]

Zhang, X.Y., Li, X., & Lin, X. (2014). The data mining technology of particle swarm optimization algorithm in earthquake prediction. Adv. *Mater. Res*., 989–994, 1570–1573. [CrossRef]

Zheng Li, Shen Chao, Tang Liang, Zeng Chunqiu, Li Tao, Luis Steven, Chen S-C. (2013). Data mining meets the needs of disaster *information management. IEEE Trans Human–Mach Syst* 2013;43(5):451–64.

Zhijun, Sun Xue Lei, Xu Yangming, Wang Zheng (2012). Review of Deep Learning. *Application Research of Computers; 2012, 29(8), p. 2806-2810.*

Zhou, F., Huang, J., Sun, B., Wen, G., & Tian, Y. (2019). Intelligent Identification Method for Natural Disasters along Transmission Lines Based on Inter-Frame Difference and Regional Convolution Neural Network. *In Proceedings of the 2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud*

Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, China, 16–18; pp. 218–222.

Appendix A
Source Codes

- Spatial data mining

```
#Importing all the required libraries for this project
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.datasets import load_breast_cancer   #Using the dataset via scikitlearn.da
tasets
from sklearn.model_selection import train_test_split

from bs4 import BeautifulSoup
from urllib import request
import pandas as pd
from shapely.geometry import Point
import geopandas as gpd
import matplotlib.pyplot as plt
import numpy as np
import ipywidgets as widgets
from IPython.display import display, clear_output
from mpl_toolkits.axes_grid1 import make_axes_locatable
from IPython.display import display, clear_output
import warnings
import seaborn as sns
import requests

%matplotlib inline


# reading in data
wikipedia = 'https://en.wikipedia.org/wiki/List_of_deadly_earthquakes_since_1900'
handler = request.urlopen(wikipedia)
soup = BeautifulSoup(handler.read(), 'html.parser')
table = soup.find_all(class_='wikitable')[0]
df = pd.read_html(str(table), header=0)[0]

pattern = r'\(.*\)|\[.*\]|\D'
df['Other Source Deaths'] = df['Other Source Deaths'].str.replace(pattern, "")
df['EM-DAT Total Deaths'] = df['EM-DAT Total Deaths'].str.replace(pattern, "")
```

```
# replacing magnitude with number
pattern = r'([^\.\d]).*'
df['Magnitude'] = df['Magnitude'].str.replace(pattern, '')

# replacing unclean longs and lats
pattern = r'\?\??'
df['Long'] = df['Long'].replace(pattern, np.nan, regex=True)
df['Lat'] = df['Lat'].replace(pattern, np.nan, regex=True)
df['Depth (km)'] = df['Depth (km)'].replace(pattern, np.nan, regex=True)


# converting to numericdf['Long'] = pd.to_numeric(df['Long'])
df['Lat'] = pd.to_numeric(df['Lat'])
df['Long'] = pd.to_numeric(df['Long'])
df['Magnitude'] = pd.to_numeric(df['Magnitude'])
df['Other Source Deaths'] = pd.to_numeric(df['Other Source Deaths'])
df['PDE Total Deaths'] = pd.to_numeric(df['PDE Total Deaths'])
df['Utsu Total Deaths'] = pd.to_numeric(df['Utsu Total Deaths'])
df['EM-DAT Total Deaths'] = pd.to_numeric(df['EM-DAT Total Deaths'])
df['Depth (km)'] = pd.to_numeric(df['Depth (km)'])


# creating new deaths column with max
df['Deaths'] = df[['PDE Total Deaths', 'Utsu Total Deaths', 'EM-DAT Total Deaths',
'Other Source Deaths']].max(axis=1)
df.columns.values[1] = 'Country'

# cleaning countries
pattern = r'(, ?| ?\().*'
df['Country'] = df['Country'].str.replace(pattern, '')
## repeated below to re-display graph ##
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
geometry = [Point(x) for x in zip(df["Long"], df["Lat"])]
gdf = gpd.GeoDataFrame(df, geometry=geometry)

fname = "PB2002_steps.json"
boundaries = gpd.read_file(fname)

countries = sorted(list(set(gdf.Country)))
countries.insert(0, 'ALL')
countries.remove("")

ax = world.plot(color='black', edgecolor='white', figsize=(25,55))
```

```
ax.axis('off')
boundaries.plot(ax=ax, color='blue', markersize=15)
gdf.plot(ax=ax, column='Magnitude', alpha=0.4, cmap='Wistia');
fig = ax.get_figure()
fig.patch.set_facecolor('#e4f4ff')
divider = make_axes_locatable(ax)
cax = divider.append_axes('right', size='3%', pad=0.01)
sm = plt.cm.ScalarMappable(cmap='Wistia', norm=plt.Normalize(vmin=2.0,
vmax=10.0))
sm._A = []
fig.colorbar(sm,cax=cax)
## repeated below to re-display graph ##

country = 'ALL'
minimum = 2.0
maximum = 9.5
world_copy = world.copy()
limits = False

m = widgets.FloatRangeSlider(
    value=(0, 15),
    min=gdf['Magnitude'].min(),
    max=gdf['Magnitude'].max(),
    description='Magnitude',
    continuous_update=False,
)

countries = sorted(list(set(gdf.Country)))
countries.insert(0, 'ALL')
countries.remove("")

d = widgets.Dropdown(
    options=countries,
    value='ALL',
    description='Country',
    disabled=False,
)

z = widgets.HBox([m,d])
display(z)

def response(change):
    try:
        clear_output(wait=True)
```

```
        global country
        global maximum
        global minimum
        global world_copy
        global limits
        if len(change['new']) == 2:
            maximum = change['new'][1]
            minimum = change['new'][0]
        else:
            country = change['new']
        display(z)
        if country.upper() == 'ALL':
            display_df = gdf[(gdf["Magnitude"] >= minimum) & (gdf['Magnitude'] <=
maximum)]
            world_copy = world
            limits = False
        else:
            display_df = gdf[(gdf["Magnitude"] >= minimum) & (gdf['Magnitude'] <=
maximum) & (gdf['Country'].str.upper() == country.upper())]
            world_copy = world[world['name'].str.upper() == country.upper()]
            xmin = display_df['Long'].min()
            xmax = display_df['Long'].max()
            ymin = display_df['Lat'].min()
            ymax = display_df['Lat'].max()
            limits = True

        ax = world_copy.plot(color='black', edgecolor='white', figsize=(25,55))
        ax.axis('off')
        boundaries.plot(ax=ax, color='blue', markersize=8)
        if limits:
            plt.xlim(xmin - 2.0, xmax + 2.0)
            plt.ylim(ymin - 2.0, ymax + 2.0)
        display_df.plot(ax=ax, column='Magnitude', alpha=0.7, cmap='Wistia');
        fig = ax.get_figure()
        fig.patch.set_facecolor('#e4f4ff')
        divider = make_axes_locatable(ax)
        cax = divider.append_axes('right', size='3%', pad=0.01)
        sm = plt.cm.ScalarMappable(cmap='Wistia',
norm=plt.Normalize(vmin=minimum, vmax=maximum))
        sm._A = []
        fig.colorbar(sm,cax=cax)
    except:
        print("Ain't no earthgdf of this magnitude in this country... Try again")
        warnings.filterwarnings('ignore')
```

```
m.observe(response, names='value')
d.observe(response, names='value')
total = len(df)
n, bins, patch = plt.hist(df["Magnitude"], histtype = 'step', range=(2.0,10.0), bins = 8)
plt.xlabel("Earthquake Magnitudes")
plt.ylabel("Frequency")
plt.title("Frequency by Magnitude")
histo = pd.DataFrame()
for i in range(0, len(n)):
    mag = str(bins[i]) + "-" + str(bins[i+1])
    freq = n[i]
    percentage = round((n[i]/total) * 100, 4)
    histo = histo.append(pd.Series([mag, freq, percentage]), ignore_index=True)

histo.columns = ['Range of Magnitude', 'Frequency', 'Percentage']
histo
warnings.filterwarnings('ignore')

gdf['Origin (UTC)'] = pd.to_datetime(gdf["Origin (UTC)"])
gdf['Year'] = gdf['Origin (UTC)'].dt.year
years = sorted(list(set(gdf['Year'])))
years.reverse()

yearly = gdf.loc[gdf['Year'] == years[0]]

y = widgets.Dropdown(
    options=years,
    value=years[0],
    description='Year',
    disabled=False)

display(y)

plt.xticks(rotation=40)
plt.title('Yearly Deaths And Proportionate Magnitude')
plt.xlabel('Country')
plt.ylabel('Deaths')
plt.scatter(x=yearly['Country'], y=yearly['Deaths'], marker='*', alpha=0.3,
s=yearly['Magnitude']**3)
plt.show()
print(yearly.loc[:, ['Country', 'Lat', 'Long', 'Magnitude', 'Deaths']])

def scatter_response(year_change):
```

```
    clear_output(wait=True)
    year = year_change['new']
    yearly = gdf.loc[gdf['Year'] == year]
    display(y)
    plt.xticks(rotation=40)
    plt.title('Yearly Deaths And Proportionate Magnitude')
    plt.xlabel('Country')
    plt.ylabel('Deaths')
    plt.scatter(x=yearly['Country'], y=yearly['Deaths'], marker='*', alpha=0.3,
s=yearly['Magnitude']**3)
    plt.show()
    print(yearly.loc[:, ['Country', 'Lat', 'Long', 'Magnitude', 'Deaths']])

y.observe(scatter_response, names='value')

c_gdf = gdf.copy()

c = widgets.Dropdown(options=countries, value='ALL', description='Country',
disabled=False,)

display(c)

plt.xticks(rotation=40)
plt.title('Country Deaths And Magnitude')
plt.ylabel('Deaths')
plt.xlabel('Magnitude')
plt.scatter(x=c_gdf['Magnitude'], y=c_gdf['Deaths'], marker='o', s=15)
plt.show()

def magnitude_response(country_change):
    clear_output(wait=True)
    country = country_change['new']
    if country.upper() == 'ALL':
        c_gdf = gdf.copy()
    else:
        c_gdf = gdf.loc[gdf['Country'].str.upper() == country.upper()]
    display(c)
    plt.xticks(rotation=40)
    plt.title('Country Deaths And Proportionate Magnitude')
    plt.ylabel('Deaths')
    plt.xlabel('Magnitude')
    plt.scatter(x=c_gdf['Magnitude'], y=c_gdf['Deaths'], marker='o', s=15)
    plt.show()
```

```
c.observe(magnitude_response, names='value')

#correlation btw depth and magnitude of earthquake
shallow = len(df[df['Depth (km)'] < 70])#18660
print(str(shallow) + " shallow earthquakes.")
intermediate = len(df[(df['Depth (km)'] > 70) & (df['Depth (km)'] < 300)]) ##3390
print(str(intermediate) + " intermediate earthquakes.")
deep = len(df[df['Depth (km)'] > 300]) #1326
print(str(deep) + " deep earthquakes.")
print(str(total) + " total earthquakes.")
print()
print(str(round(shallow/float(total) * 100, 2)) + " % of earthquakes are shallow.")
print(str(round(intermediate/float(total) * 100, 2)) + " % of earthquakes are
intermediate.")
print(str(round(deep/float(total) * 100, 2)) + " % of earthquakes are deep.")
1158 shallow earthquakes.
83 intermediate earthquakes.
5 deep earthquakes.
1340 total earthquakes.

86.42 % of earthquakes are shallow.
6.19 % of earthquakes are intermediate.
0.37 % of earthquakes are deep.

mag_dep = gdf.groupby(['Country','Magnitude','Depth (km)'])['Deaths'].sum()
mag_dep = mag_dep.reset_index()
mag_dep
plt.title('Magnitude to Depth (km) correlation')
plt.xlabel('Magnitude')
plt.ylabel('Depth (km)')
plt.scatter(mag_dep['Magnitude'], mag_dep['Depth (km)'], marker='*', alpha=0.3)#,
s=mag_dep['Deaths'])

d_c = widgets.Dropdown(
    options=countries,
    value='ALL',
    description='Country',
    disabled=False,
)
display(d_c)

def plot_d(c_change):
    clear_output()
    country = c_change['new']
```

```
    plot_mag_dep = mag_dep.copy()
    test_mag_dep = mag_dep.loc[mag_dep['Country'] == country]
    display(d_c)
    plt.title('Magnitude to Depth (km) correlation for ' + country)
    plt.xlabel('Magnitude')
    plt.ylabel('Depth (km)')
    plt.scatter(test_mag_dep['Magnitude'],test_mag_dep['Depth (km)'], marker='*',
alpha=0.3)

d_c.observe(plot_d, names='value')

c_y_d = gdf.groupby(['Country','Year','Magnitude'])['Deaths'].sum()
c_y_d=c_y_d.reset_index()
c_y_d_select = c_y_d.loc[c_y_d['Country'] == 'Colombia']


plt.title('Magnitude by year per country, sized by deaths')
plt.xlabel('Year')
plt.xticks(rotation='vertical')
plt.ylabel('Magnitude')
plt.scatter(c_y_d_select['Year'],c_y_d_select['Magnitude'], marker='*', alpha=0.3,
s=c_y_d_select['Deaths'])

countries.remove('ALL')

d_y_d = widgets.Dropdown(
    options=countries,
    value='Colombia',
    description='Country',
    disabled=False,
)
display(d_y_d)

def plot_c_y_d(c_y_d_change):
    clear_output()
    country = c_y_d_change['new']
    c_y_d_select = c_y_d.loc[c_y_d['Country']==country]
    display(d_y_d)
    plt.title('Magnitude by year for ' + country + ' sized by deaths')
    plt.xlabel('Year')
    plt.xticks(rotation='vertical')
    plt.ylabel('Magnitude')
    plt.scatter(c_y_d_select['Year'],c_y_d_select['Magnitude'], marker='*', alpha=0.3,
s=c_y_d_select['Deaths'])
```

```
d_y_d.observe(plot_c_y_d, names='value')


#Using ML Model


from sklearn.svm import SVC
from sklearn.metrics import classification_report , confusion_matrix
from sklearn.svm import SVC
svm_model = SVC(kernel = 'linear', random_state = 0)
svm_model.fit(x_train, y_train)
y_predict =svm_model.predict(x_test)
cm = confusion_matrix(y_test,y_predict)

#model improvisation
min_train =x_train.min()
range_train =(x_train - min_train).max()
x_train_scaled =(x_train-min_train)/range_train

from sklearn.metrics import f1_score
f1_score(y_test, yhat, average='weighted')

from sklearn.metrics import jaccard_similarity_score
jaccard_similarity_score(y_test, yhat)

from sklearn.cluster import KMeans

        km=KMeans(n_clusters=cluster).fit(final)
        labels=km.labels_
        centroids=km.cluster_centers_
        c_mean_distances = []
        for i, (cx, cy) in enumerate(centroids):
                mean_distance = k_mean_distance(final, cx, cy, i, labels)
                c_mean_distances.append(mean_distance)
        len_centroid=len(centroids)
        length={len(final[np.where(labels == i)]) for i in range(km.n_clusters)}
        points={i: final[np.where(labels == i)] for i in range(km.n_clusters)}
        plt.scatter(final[:,0],final[:,1], c=[plt.cm.nipy_spectral(float(i)/10) for i in
labels])
        plt.scatter(centroids[:,0],centroids[:,1],marker='x')
        plt.savefig("visualization.png")
        plt.clf()
        fileinfo=base64.b64encode(readbytesfile("visualization.png")).decode()
```

```
        end=time.time()
        elapsed=end-start
        distance=[]
        fields=zip(c_mean_distances,length,centroids)
        return
render_template("main_page.html",fileinfo="data:image/jpg;base64,"+fileinfo,timeela
psed=str(elapsed),length=length,points=str(points),distance=c_mean_distances,fields=
fields)

        #return '''<html><body><img src="data:image/jpg;base64,'''+fileinfo+'''"
style="width:500px;height:500px;"/><br>Time
Elapsed:'''+str(elapsed)+'''<br>'''+str(length)+'''<br>'''+str(points)+'''</body></html>'''

        start=time.time()
        sql="select sum(case when survived=1 and sex='female'  then 1 end) as
first,sum(case when survived=0 and sex='female'  then 1 end) as second from titanic";
        cursor=db.cursor();
        cursor.execute(sql);
        result=cursor.fetchall()
        val=[]
        for vals in result[0]:
                val.append(float(vals))
        labels=["survived","not survived"]
        total=0
        for vals in val:
                total=total+vals
        legendtab=[]
        for i in range(len(labels)):
                legendtab.append(labels[i]+" - "+str(round(val[i]/total*100,2))+"%")
        index=np.arange(len(labels))
        patch=plt.pie(val,shadow=False,startangle=140,pctdistance=0.9, radius=2)
        plt.legend(patch[0],legendtab,bbox_to_anchor=(0.5,0.5),bbox_transform=plt.g
cf().transFigure,loc="best")
        plt.clf()
        fileinfo=base64.b64encode(readbytesfile("visualization.png")).decode()
        end=time.time()
        elapsed=end-start
        elapsed="%.3f" % elapsed
        return
render_template("main_page.html",piefileinfo="data:image/jpg;base64,"+fileinfo,pieti
me=str(elapsed)+"s")
```

Appendix B

Ethical Approval Document

NEAR EAST UNIVERSITY

ETHICAL APPROVAL DOCUMENT

Date: 30/04/2022

To the Institute of Graduate Studies;

For the thesis project entitled "INVESTIGATION AND FORECAST OF COMMON CATASTROPHE UTILIZING SPATIAL INFORMATION MINING PROCEDURE," the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

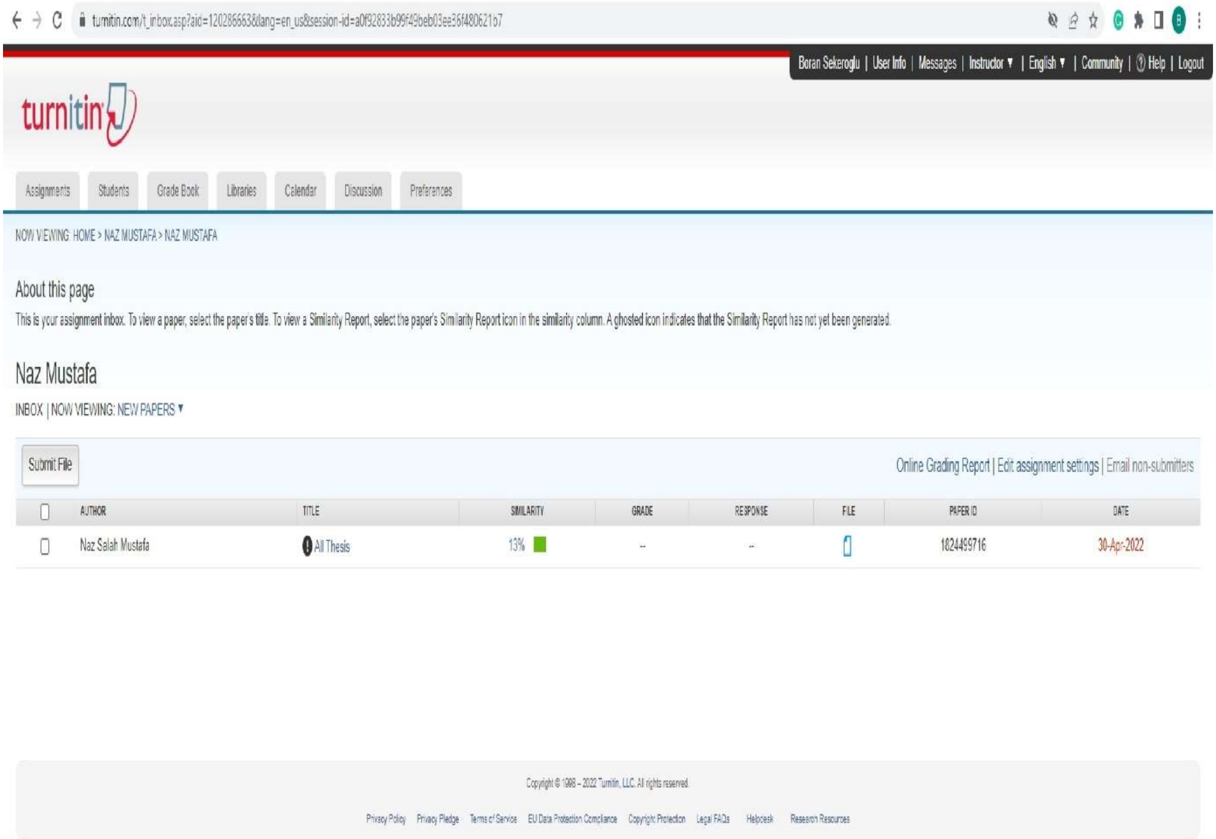
Title: Assoc. Prof. Dr.

Name Surname: Boran Şekeroğlu

Signature:

Role in the Research Project: Supervisor

Appendix C
Similarity Report



Assoc.Prof. Dr. Boran Şekeroğlu
Supervisor