An Experiment on Parameter Selection for Landslide Susceptibility Mapping using TF-IDF

Sonam Lhamu Bhutia\textsuperscript{1}, Samarjeet Borah\textsuperscript{2}, Ratika Pradhan\textsuperscript{3} and Bhushan Sharma\textsuperscript{4}

\textsuperscript{1,2,3,4}Sikkim Manipal Institute of Technology, Majhitar, Sikkim, India-737136

samarjeet.b@smit.smu.edu.in

Abstract. Landslide can be considered as one of the most common natural threats faced mostly by the people living in the hilly and mountainous regions. Every year, mostly during the monsoons, landslide disrupts the lives of people living in these areas, at times it also leads to the damage of lives, properties etc. Landslide susceptibility mapping has become an essential measure for prevention of losses due to landslide. In this paper, some input factors that are most common for landslide susceptibility map pertaining to the hilly and mountainous regions have been identified using TF-IDF method.

Keywords. Landslide, Landslide Susceptibility Map (LSM), Parameter, Parameter Selection, TF-IDF

1. Introduction

One can define landslide as the movement of earth, debris and rock which is driven by a gravitational force down in a slope [1]. There can be several factors that causes landslides, these factors may be geological, physical, morphological or man-made where in these factors are always triggered by a triggering factor. Landslides have always been a major area of concern in the hilly and mountainous regions as it causes heavy damage to both lives and property [1]. This has led to the need for development of various models to assess the vulnerability of the land to the landslides and thus zone these areas according to its degree of vulnerability and this can be done using landslide susceptibility map.

Landslide susceptibility map classifies those areas with most likely to occur a landslide, where segregation is done from low to high. These maps take into account those areas that are most likely to occur [2]. Thus, one of the key areas of research today is landslide susceptibility mapping (LSM) and it has become one of the vital tools in decision making processes pertaining to development plan and strategies. The most significant phase in LSM is determining all vital input features for landslide, for any given study area [25]. Researchers have divided these factors affecting the landslide into conditioning factors and triggering factors. There is not just single conditioning factor that affects the landslides but several, however, there is only one single triggering factor.

For LSM of any study area, landslide conditioning factor plays a very significant and prominent role in determining its accuracy [28]. Better is accurateness of any landslide susceptibility map when there is better selection of the causative factors [27]. Although conditioning factors form an integral portion of the landslide susceptibility mapping, it is also crucial to optimally select these factors as inclusion of...
more numbers of condition factors may lead to elimination of crucial factors. Also these conditioning factors varies according to the study area.

Selection of the conditioning factors for the landslide is either done according to the experts’ opinion or as per the previous existing literature or both. These parameters may vary according to the study area. However, within these parameters, some of them are common for all the study area pertaining to the regions that are hilly and mountainous. Hence, the need for efficient selection of these landslide causative factors has received considerable attention in these years.

In view of this, this study aims to identify and select some of the most common used parameters (optimal landslide conditioning factors). In addition to these parameters, a few may be considered depending on the study area. The proposed work will try to include the most essential parameters and exclude the least significant one based on the machine learning technique. This manuscript is structured as follows – Some influential works from literature are discussed in Section 2, section 3 highlights proposed methodology and implementation, Results are highlighted in Section 4 and the work has been concluded in the fifth section.

2. Literature Review
This section contains few works from literature on landslide susceptibility mapping. Here, three general research works on landslide susceptibility mapping is followed by works that employ Term Frequency-Inverse Document Frequency (TF-IDF) as standard technique.

A decision tree based technique for the same can be found in [45] which employs ensemble learning algorithms known as decision forest. Here, conditioning factors included were slope length, land-use, sediment transport index (STI), normalized-difference-vegetation-index (NDVI), aspect, topographical-roughness-index (TRI), stream-power-index (SPI), drainage-density (DD), topographic-wetness-index (TWI) elevation, slope, lithology etc. However, all these conditioning factors that were considered were in correlation with the study related to the distribution of existing landslides. The proposed work will try to take care of irrelevant factors and select these parameters which are common to all the study areas having the same landscape. Ashournejad. Q et al. [46], proposed a model that was based on both knowledge-based as well as data focused approaches. For knowledge based approach, fuzzy gamma operator a simple additive weighting is used. Probabilistic neural network and radial basis function link network methods are used for data centric approach. The input parameters that were considered for the model were fault density, geology, distance to roads, rainfall, DD, slope, aspect, distance-to-faults, landuse, drainage, soil, distance-to-river. Here, the input parameters for the model have been considered from the existing literature and experts’ opinion. The proposed work will try to extract some of the most common parameters using machine learning technique. Arishma Gadtaula and Subodh Dhakal [47], proposed a technique to model the landslide susceptibility map through bivariate statistical model, Weight-of-Evidence (WoE). The input parameters that were considered were land-use, elevation, aspect, geology slope, peak-ground-acceleration (PGA), curvature (plan and profile). These conditioning factors were taken into account based on experts ‘knowledge; these factors were chosen based on their relevance to the study area.

2.1. Information Retrieval using TF-IDF
TF-IDF is an information retrieval algorithm [19] that uses a statistical measure to estimate the relevance of word to its document within a collection of document known as corpus. It is an important technique which is primarily used for text analysis that checks the relevance of a word to its documents within a corpus. Some of its important applications are keyword extraction and information retrieval [20]. Here, two metrics are taken, the first one is frequency at which a word appears in a document, it is known as term-frequency(TF) and the second one is the inverse document of that word within the corpus known as inverse document frequency(IDF). Then the multiplication of these two metrics will result in TF-IDF score of a word in a document [20]. This technique does not measure the frequency of the word as weights are assigned to the word and hence it measures the relevance of words in corpus. the calculated TF-IDF scores replaces the word count, in a corpus. TF-IDF technique works on the principle that lower
the TF-IDF score higher is the relevance of a word and higher the score of TF-IDF lesser is its relevance [21].

The mathematical representation of the same can be given as:

$$TF-IDF(T,D) = TF(T,D) \times \log \frac{N}{DF+1}$$  \[23\]

Where,

- $T = \text{term(word)}$
- $D = \text{document}$
- $N = \text{total number of documents}$ [22]
- $TF = \text{term frequency}$
- $DF = \text{document frequency}$

In this study, taking the application of keyword extraction into consideration, some of the most used parameters can be extracted from a corpus which can be included in the list of most essential parameters for landslide susceptibility mapping. Before going to the proposed methodology, few influential research works using TF-IDF are reviewed below:

Ammar Ismael Kadhim[48], made a study on extraction of various features on Twitter. A comparative analysis for two techniques viz. BM25 and TF-IDF was done where TF-IDF out performed BM25 in terms of performance classification metrics. Both these techniques were used to extract the keywords from the data that was used in Twitter. TF-IDF technique was found to be more stable.

Jinghuan Guo, et al. [49], proposed a strategy based on TF-IDF for feature selection for recognition of various activities in smart homes. The proposed strategy has used statistical information for individual activity recognition. TF-IDF is shown to be an efficient technique in this case.

Manal Mohammed and Nazlia Omar [50], proposed a technique for automatic classification of the questions using feature extraction based on modified TF-IDF and word2vec which were further classified using the ML classification algorithms. The classification of questions was centered on cognitive domain of Bloom’s Taxonomy. Ignacio Arroyo-Fernández, et al. [51], proposed a technique known as Word Information Series for Sentence Embedding (WISSE) for representing a sentence as word information series. The proposed technique involves the usage of TF-IDF. The proposed technique, WISSE was better as compared to other existing techniques.

As TF-IDF have been proposed for feature extraction, hence, in this work, an effort has been made to apply the TF-IDF technique for extraction of the most common parameters for a common type of terrain to model of landslide susceptibility map. Addition of some other parameters in conjunction to these common parameters may help in ensuring that all the essential parameters for modelling of the landslide susceptibility map is included.

3. Landslide Susceptibility Mapping using TF-IDF - Proposed Methodology and Implementation

The proposed aims to extract of some of the most common parameters may be considered in modelling of landslide susceptibility map. The data set considered here is consisting of 52 PDF files, which are basically research papers from the domain of landslide susceptibility mapping published in various journals and conferences. These are carefully selected based on their accuracy in prediction. In the experiment parameters are extracted based on the TF-IDF score. The parameters with lesser TF-IDF score are taken into consideration. All the research papers from the dataset are given in the Table 1 along with the parameters or attributes used and accuracy in prediction.
Table 1. Description of Input Dataset

| Author and Year | Attributes Used                                                                 | Prediction Accuracy % |
|-----------------|---------------------------------------------------------------------------------|----------------------|
| Saro Lee [2003] | Timber diameter, soil effective thickness Slope, rainfall, soil-drainage, timber age, Curvature, Soil texture | 87.78                |
| Ayalew, L, et al. [2004] | Intercept, slope-angle, slope-gradient, proximity to roads, aspect, bed-rock-slope, Lithology, lineaments, elevation | 83.58                |
| Nagarajan R, et al. [2000] | daily rainfall, slope angle, rock stratification, slope facet, relieve relief, rainfall intensity, weathering, lineament, soil type, soil depth, land cover, soil rock interface, drainage density | 72.2                 |
| Hossein Moayedi, et al. [2018] | Elevation, slope degree, distance to river, lithology, twi, slope aspect, stream-power-index, soil types, distance-to-road, land-use, distance-to-fault, curvature. | 97.17                |
| Baeza C, et al. [2001] | Soil-type, altitude, slope-aspect, slope-angle, tree density, slope complexity, length of water shed land cover, thickness of superficial deposits | 88.8                 |
| Ge Yan, et al. [2018] | topographic roughness index, proximity to streams, topographic wetness index, elevation, stream power index, slope, proximity to roads, sediment transport index measurements curvature lithology, altitude, aspect | 87.51                |
| Yang Yang, et al. [2019] | Slope, land use, inhabited areas to the mapping unit distance, rock group, surface roughness of the terrain, digital-elevation-model (DEM), distance-to-road | 86.09                |
| Cheng Zhong, et al. [2019] | Distant-view-with-high-spatial, temporal, and spectral-resolutions | 78.26                |
| Van Westen et al. [2003] | LULC, rainfall, TWI, soil-texture, aspect, slope, geology, distance-from-road, distance-from-lineament, altitude, distance-from-river, NDVI | 76                   |
| Jagabandhu Roy, et al. [2019] | Lineament, slope-density, curvature, LULC, drainage-buffer, aspect, lithology | 90                   |
| Swati Sharma, et al. [2018] | slope angle, rainfall, elevation, LULC, road-density, drainage-density, curvature(profile), geology, NDVI, slope-aspect | 87.12                |
| Subodh Chandra Pal, et al. [2019] | drainage, slope, earthquake, land use, geomorphology, rainfall, soil, aspect and relief, lineament, lithology | 89                   |
| Amit Chawla, et al. [2019] | elevation, rainfall, LULC, distance-to-lineament, geology, lineament-density, TWI, curvature, topographic wetness, NDVI, slope-aspect, soil, drainage-density, distance-to-drainage, SPI | 92.3                 |
| Subrata Mondal, et al. [2018] | Distance-to-drainage, aspect, curvature(plan), sub-watershed-basin, road-density, distance-to-faults, drainage-density, fault-density, elevation geology, SPI, curvature(profile), slope-angle, distance-to-roads, slope-length, surface-area-ratio, TWI, distance-to-ridges, LULC | 78.2                 |
| Yesilnacar E, et al. [2005] | Curvature(plan), distance-from-drainage, slope-angle, distance-from-lineament, aspect, geology, LULC, distance-from-river | 89                   |
| Lee S, et al. [2006] | aspect, profile curvature, lithology, plan curvature, slope gradient, elevation | 86.37                |
| Ayalew L, et al. [2004] | TWI, distance from roads, aspect, distance-from-faults, lithology, SPI, distance-from-rivers, altitude, curvature(plan), slope-degree, LULC | 70                   |
| Pourghasemi H R, et al. [2012] | lithology, road density, aspect, profile-curvature, relative-relief, slope length, TWI, slope curvature, distance from road, TPI, elevation, catchment area, slope degree, distance from rivers, planar curvature, SPI | 89.7                 |
| Braun A, et al. [2018] | Slope-length, distance-from-river, aspect, slope, curvature(plan), NDVI, TWI, distance-from-faults, convergence-index, elevation, lithology, LULC, rainfall, soil-type, distance-from-roads | 91.2                 |
| Authors          | Contributions                                                                 | Impact Factor |
|------------------|-------------------------------------------------------------------------------|---------------|
| Chen W, et al.   | STI, NDVI, curvature(plan), distance-to-roads, TWI, altitude, slope angle, distance-to-faults, curvature(plan), slope-degree, LULC, distance-to-rivers, SPI | 83.9          |
| Arabameri A, et al. [2019] | LULC, SPI, drainage-density, elevation, distance-from-river, rainfall, convergence-index, slope-aspect, distance-from-fault, curvature(plan), slope-length, lithology, distance-from-road, slope, NDVI | 82            |
| Chen W, et al.   [2019] | Slope-angle, TWI, elevation, STI, land use, SPI, NDVI, slope-aspect, curvature(profile) | 89            |
| Can. A, et al.   [2007] | slope, WTI, aspect, elevation, NDVI, lithology, rainfall | 81.7          |
| Kutlug Sahin E, et al.[2019] | Elevation, aspect, STI, TWI, TRI, drainage density, NDVI, slope, lithology, SPI, slope length, LULC | 83.3          |
| Kornejady. A., et al.[2018] | TWI, slope, soil texture, plan curvature distance to road, profile curvature, altitude, rainfall, distance to stream | 83.1          |
| Chen. W, et al.  [2018] | Slope angle, profile-curvature, NDVI, distance-to-roads, elevation, curvature(plan), lithology, LULC, slope-aspect, distance-to-rivers, rainfall, distance-to-faults. | 84.7          |
| Gadtaula A, et al. [2018] | Curvature(profile), lithology, peak-ground-acceleration(PGA), profile curvature, elevation, slope | 80            |
| Nohani, et al.   [2019] | Slope-aspect, elevation, lithology, NDVI, distance-from-river, slope-angle, distance-from-fault, curvature(profile), LULC, distance-from-road, curvature(plan) | 84            |
| Kavzoglu T, et al. [2018] | TWI, elevation, NDVI, LULC, lithology, slope-aspect, curvature(plan) | 87.23         |
| Pham B T, et al. [2018] | TRI, valley-depth, TWI, lithology, elevation, slope-aspect, distance-to-roads, soil-type, distance-to-river, curvature(profile), slope-length, STI slope-angle, LULC, SPI, distance-to-lineaments | 96.6          |
| Yan F, et al.    [2018] | rainfall, slope-aspect, distance-to-road, slope-angle, NDVI, distance-to-river, lithology, distance-to-construction, distance-to-fault | 88.6          |
| Dou J, et al.    [2019] | Drainage, lithology, elevation, curvature(plan), slope-aspect, density-of-geologic-boundaries | 87            |
| Boualla O, et al. [2017] | slope, lithology, topography, geology, hydrology, rainfall, distance-to-roads, elevation, slope-gradient | 79.16         |
| Ashournejad Q, et al. [2019] | Drainage-density, distance-to-faults, rainfall, distance-to-roads, aspect, LULC, fault-density, soil, slope, geology, distance-to-river | 83.5          |
| Pham B T, et al. [2018] | slope, distance to rivers, curvature, SFM, distance-to-roads, aspect, LULC valley-depth, geomorphology | 79.19         |
| Juliev M, et al. [2018] | Soil, distance-to-lineaments, distance-to-streams, geology, distance to roads, distance-to-faults, slope-aspect, LULC, elevation, slope-angle | 80            |
| Peethambaran B, et al. [2018] | Curvature(profile), slope-gradient, TWI, slope-aspect, altitude, plan curvature | 84.4          |
| Shahab H, et al. [2015] | lithology, distance-to-faults, precipitation, distance-to-drainage, slope-aspect, LULC, soil, NDVI, distance-to-roads | 96            |
| Pham B T, et al. [2017] | Curvature(profile), LULC, rainfall, distance-to-roads, elevation, lineament density, road-density, profile(curvature), soil-type, distance-to-lineaments, slope-angle, river-density, distance-to-rivers, slope-aspect | 88.1          |
| Tsangaratos P, et al. [2016] | lithology, slope-angle, distance-from-geological-boundaries, elevation, slope-aspect, distance-from-hydrographic-network, distance-from-road | 83.97         |
| Kavzoglu T, et al. [2015] | lithology, slope-length, elevation, NDVI, aspect, TWI, soil-depth, distance-to-lineaments, distance-to-drainage, curvature(plan), TPI, slope, LULC, curvature(plan), distance-to-road | 90.23         |
| Chen W, et al. [2017] | Lithology, slope-aspect, altitude, distance-to-drivers, NDVI, profile-curvature, slope-angle, distance-to-roads, SPI, slope-aspect | 87.51         |
| Kutlug Sahin E, et al. [2019] | Lithology, slope-aspect, distance-to-drivers, NDVI, profile-curvature, slope-angle, distance-to-roads, SPI, slope-aspect | 90.02         |
Algorithm

**Input:** PDF File, Size Count 52 //Generated based on accuracy of the technique discussed in the research paper

**Output:** A Set of Parameters

**Step 1** Calculate Term Frequency (TF) //Measures the frequency of occurrence of a term in a document

\[
TF = \frac{\text{Number of Frequency of Occurrence of A Term in A Document}}{\text{Total Number of Terms in the Document}}
\]

**Step 2** Calculate Document Frequency (DF) //Frequency of the term that has occurred in a document

**Step 3** Calculate Inverse Document Frequency (IDF) //Measure the significance of term in a document

\[
IDF(T) = \log \frac{N}{(DF+1)}
\]

**Step 4** Calculate the TF-IDF score

\[
TF-IDF(T, D) = TF(T, D) \times \log \frac{N}{(DF+1)}
\]

**Step 5** Select the final parameters

---

The experiment has been conducted on an Intel i5 machine with 16 GB of RAM and 4 GB graphics card. It is implemented using Python 3.6 along with IDE for python as IDLE. Here PDF files are extracted using Pypdf4 and Xlrd is used to read keywords from .xlsx file.

4. Result and Discussion

The result is obtained after application of TF-IDF technique for a given set of input in Table 1, the word section implies parameters. These are the common parameters, that have been selected as per their TF-
IDF score. The parameter having the lowest TF-IDF score indicates that a parameter for the LSM has been frequently used in most of the pdfs used which implies that it is most used one.

One of the key advantage of TF-IDF is, that it helps filter out unwanted words from the document and these unwanted words are known as stop words [20].

Table 2. Results for TF-IDF Technique

| SN  | Words                                      | TF-IDF Score        |
|-----|--------------------------------------------|---------------------|
| 1   | Topographical Roughness Index              | 0.517857047         |
| 2   | Precipitation                              | 0.524361544         |
| 3   | Sediment Transport Index                   | 0.536671365         |
| 4   | Land Use And Land Cover                    | 0.552833374         |
| 5   | Altitude                                   | 0.554484993         |
| 6   | Erosion                                    | 0.559493921         |
| 7   | Slope Angle                                | 0.5726017           |
| 8   | Stream Power Index                         | 0.58556954          |
| 9   | Normalized Diff. Vegetation Index          | 0.625228377         |
| 10  | Topographic Wetness Index                  | 0.625310964         |
| 11  | Geology                                    | 0.672952719         |
| 12  | Drainage                                   | 0.68641605          |
| 13  | Density                                    | 0.707580347         |
| 14  | Rainfall                                   | 0.731869218         |
| 15  | Curvature                                  | 0.753358            |
| 16  | Elevation                                  | 0.77674             |
| 17  | Soil                                       | 0.840404            |
| 18  | Aspect                                     | 0.899644            |

5. Conclusion

Of the several parameters taken into consideration, there are some of the most commonly used parameters. Here, TF-IDF technique have been used to identify some most commonly used parameters that may be included for improvement of model for landslide susceptibility map pertaining to mountainous and hilly mountainous regions. These parameters are based on their TF-IDF scores obtained. Lesser TF-IDF score presents that the term is the most used one, thus implies that it is the most significant one. The use of TF-IDF technique have helped in identifying the most frequently used parameters for landslide susceptibility mapping. Thus, it indicates that these are the most significant parameters which can be considered for development of landslide susceptibility maps for mountainous regions.

The dataset is prepared locally by manually selecting the papers with good accuracy score from the same domain. Hence, extraction of the parameters along with its accuracy was a bit of a challenge. Also selection of parameters manually from PDFs for the preparation of datasheet is time consuming. The present work, has taken into account only those study areas which have taken rainfall as the triggering factor into consideration. Future work may include other triggering factors that causes landslides.

References
[1] Bhutia S.L., Borah S., Pradhan R. (2020) Landslide Susceptibility Mapping: Development Towards a Machine Learning-Based Model. In: Sarma H., Bhuyan B., Borah S., Dutta N. (eds) Trends in Communication, Cloud, and Big Data. Lecture Notes in Networks and Systems, vol 99. Springer, Singapore.
[2] Department of Communications, C. A. (2020). Landslide-Susceptibility-Mapping.aspx. Retrieved from https://www.gsi.ie/en-ie/Pages/default.aspx:https://www.gsi.ie/en-ie/programmes-and-projects/geohazards/projects/Pages/Landslide-Susceptibility-Mapping.aspx
[3] R. Nagarajan, et al. “Landslide hazard susceptibility mapping based on terrain and climatic factors for tropical monsoon regions”, Bull Eng Geol Env (2000) 58 : 275–287 7 Q Springer-
Verlag.

[4] Hossein Moayedi, et al. “Modification of landslide susceptibility mapping using optimized PSOANNN technique”, Received: 22 July 2018 / Accepted: 17 September 2018 / Published online: 21 September 2018 Springer-Verlag London Ltd. doi: 10.1007/s00366-018-0644-0.

[5] Baeza, C., & Corominas J, “Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surface Processes and Landforms, 26(12), 1251–1263.doi:10.1002/esp.263, 2001.

[6] Ge Yan, et al. “Optimizing landslide susceptibility mapping in the Kongtong District, NW China: Comparing the subdivision criteria of factors”, Geocarto International. doi:10.1080/10106049.2018.1499816, 2018

[7] Yang Yang, et al. “Local-scale landslide susceptibility mapping using the B-GeoSVC model”, Received: 1 May 2018 Accepted: 21 March 2019 © Springer-Verlag GmbH Germany part of Springer Nature 2019, DOI 10.1007/s10346-019-01174-y

[8] Cheng Zhong, et al. “Landslide mapping with remote sensing: challenges and opportunities”, International Journal of Remote Sensing, DOI: 10.1080/01431161.2019.1672904, 2019.

[9] Van Westen, C. J., Rengers, et al. “Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment”, Natural Hazards, 30(3), 399–419.doi:10.1023/b:nhaz.000007097.42735, 2003.

[10] Jagabandhu Roy and Sunil Sah, “Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. Doi:10.1186/s40677-019-0126-8, Geoenvironmental Disasters, 2019 – Springer

[11] Swati Sharma & A. K. Mahajan, “A comparative assessment of information value, frequency ratio and analytical hierarchy process models for landslide susceptibility mapping of a Himalayan watershed, India. Doi: 10.1007/s10064-018-1259-9, Bulletin of Engineering Geology and the Environment , Springer, 2018.

[12] Subodh Chandra Pal, Indrajit Chowdhuri, “GIS-based spatial prediction of landslide susceptibility using frequency ratio model of Lachung River basin, North Sikkim, India. Doi: 10.1007/s42452-019-0422-7, Earth and Environmental Sciences, Springer, 2019

[13] Amit Chawla,Srinivas Pasupuleti, et al. “Landslide Susceptibility Zonation Mapping: A Case Study from Darjeeling District, Eastern Himalayas, India”, Doi: 10.1007/s12524-018-0916-6, Journal of the Indian Society of Remote Sensing, Springer, 2019.

[14] Subrata Mondal & Sujit Mandal, “Landslide susceptibility mapping of Darjeeling Himalaya, India using index of entropy (IOE)”, Doi:10.1007/s12518-018-0248-9, Applied Geomatics volume 11, pages129–146(2019)

[15] Yesilnacar, E., & Topal T, “Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)”, Engineering Geology, 79(3-4), 251–266.doi:10.1016/j.enggeo.2005.02.002, 2005

[16] Lee,S, et.al, “Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models”,Environmental Geology, 50(6), 847–855.doi:10.1007/s00254-006-0256-7, Springer,2006

[17] Ayalew. L, et al., “Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides, 1(1), 73–81.doi:10.1007/s10346-003-0006-9, Springer 2004.

[18] Pourghasemi, H. R., et al. “Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Natural Hazards, 63(2), 965–996.doi:10.1007/s11069-012-0217-2”, Springer, 2012.

[19] G.B,(2018b, March 6). The TF*IDF Algorithm Explained. https://Www.Onely.Com/Blog/What-Is-TF-Idf/. https://www.onely.com/.

[20] Li-Ping Jing, Hou-Kuan Huang, & Hong-Bo Shi. (2003, February 19). Improved feature selection approach TFIDF in text mining. Proceedings. International Conference on Machine Learning and Cybernetics, 944–946. https://doi.org/10.1109/icmlc.2002.1174522
[21] Feature Extraction Techniques - NLP. (2020, March 6). GeeksforGeeks. https://www.geeksforgeeks.org/feature-extraction-techniques-nlp/

[22] Hamdouy, Y. (2019, December 10). TF(Term Frequency)-IDF(Inverse Document Frequency) from scratch in python. Medium. https://towardsdatascience.com/term-frequency-idf-inverse-document-frequency-from-scratch-in-python-6c2b61b78558

[23] Braun.A, et al., “Landslide Susceptibility Mapping in Tegucigalpa, Honduras, Using Data Mining Methods”. IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018 - Volume 1, 207–215. doi:10.1007/978-3-319-93124-1_25, Springer, 2018.

[24] Hyun-JooOh, BiswajeetPradhan, “Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area”, Computers & Geosciences, Volume 37, Issue 9, Pages 1264-1276, Springer, September 2011.

[25] Wei Chen, et al., “Landslide susceptibility mapping based on GIS and support vector machine models for the Qianyang County, China”, Environmental Earth Sciences volume 75, Article number: 474 (2016), Springer.

[26] Haoyuan Hong, Seyed Amir Naghibi, Hamid Reza Pourghasemi, Biswajeet Pradhan. "GIS-based landslide spatial modeling in Ganzhou City, China", Arabian Journal of Geosciences, Article number: 112 (2016), Springer.

[27] Matebie Meten, et al., “Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia”, GeoenvIRONMENTAL DISASTERS Volume 2, Article number: 9 (2015), Springer.

[28] Arabameri. A., Pradhan. B, et al., “GIS-based landslide susceptibility mapping using numerical risk factor bivariate model and its ensemble with linear multivariate regression and boosted regression tree algorithms”, Journal of Mountain Science, 16(3), 595–618. doi:10.1007/s11629-018-5168-y, Springer, 2019.

[29] Chen. W, et al., “GIS-Based Landslide Susceptibility Evaluation Using Certainty Factor and Index of Entropy Ensembled with Alternating Decision Tree Models”. Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques, 225–251. doi:10.1007/978-3-319-73383-8_10, Springer, 2018.

[30] Arabameri.A, et al., “An Ensemble Model for Landslide Susceptibility Mapping in a Forested Area”. Geocarto International, 1–25. doi:10.1080/10106049.2019.1585484", Taylor&Francis Online, 2019.

[31] Chen. W, et al. “Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility”. CATENA, 172, 212–231. doi:10.1016/j.catena.2018.08.025, ScienceDirect, 2019.

[32] Can.A, et al., ”Landslide susceptibility mapping at Ovacik-Karabük (Turkey) using different artificial neural network models: comparison of training algorithms”. Bulletin of Engineering Geology and the Environment. doi:10.1007/s10064-017-1034-3, Springer, 2019.

[33] Kutlug Sahin.E, et al., “Performance Analysis of Advanced Decision Tree-Based Ensemble Learning Algorithms for Landslide Susceptibility Mapping”. Geocarto International, 1–23. doi:10.1080/10106049.2019.1641560, Taylor&Francis Online, 2019.

[34] Kornejady, A., Pourghasemi, H. R., & Afzali, S. F. (2018). Presentation of RFFR New Ensemble Model for Landslide Susceptibility Assessment in Iran. Advances in Natural and Technological Hazards Research, 123–143. doi:10.1007/978-3-319-77377-3_7, Springer, 2018.

[35] Chen.W, et al., “Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBF Network models for the Long County area (China)”. Bulletin of Engineering Geology and the Environment. doi:10.1007/s10064-018-1256-z, Springer, 2018.

[36] Gadtaula. A. & Dhakal. S, “Landslide susceptibility mapping using Weight of Evidence Method in Haku, Rasuwa District, Nepal”. Journal of Nepal Geological Society, 58, 163–171. doi:10.3126/jngs.v58i0.24601, 2019.
[37] Nohani, Moharrami, et al., “Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. Water, 11(7), 1402. doi:10.3390/w11071402, MDPI, 2019

[38] Kavzoglu. T, “Machine Learning Techniques in Landslide Susceptibility Mapping: A Survey and a Case Study. Advances in Natural and Technological Hazards Research, 283–301. doi:10.1007/978-3-319-77377-3_13”, Springer, 2018.

[39] Pham, B, et al. “A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling”. Bulletin of Engineering Geology and the Environment. doi:10.1007/s10064-018-1281-y, Springer, 2018.

[40] Yan, F, et al. , “A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model”. Geomorphology https://doi.org/10.1016/j.geomorph.2018.10.024, ScienceDirect, 2019.

[41] Dou, J, et al. “Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. Remote Sensing, 11(6), 638. doi:10.3390/rs11060638, MDPI, 2019

[42] Boualla. O.,et al. “GIS-based landslide susceptibility mapping in the Safi region, West Morocco. Bulletin of Engineering Geology and the Environment. doi:10.1007/s10064-017-1217-y”, Springer, 2017

[43] Ashournejad. Q, et al. “Hazard zoning for spatial planning using GIS-based landslide susceptibility assessment: a new hybrid integrated data-driven and knowledge-based model”. Arabian Journal of Geosciences, 12(4). doi:10.1007/s12517-019-4236-0, Springer, 2019

[44] Emrehan Kutlug Sahin & Ismail Colkesen (2019) Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping, Geocarto International, DOI: 10.1080/10106049.2019.1641560

[45] Ashournejad, Q., Hosseini, A., Pradhan, B. et al. (2019), Hazard zoning for spatial planning using GIS-based landslide susceptibility assessment: a new hybrid integrated data-driven and knowledge-based model. Arab J Geosci 12, 126 https://doi.org/10.1007/s12517-019-4236-0

[46] Gadtaula, Arishma & Dhakal, Subodh. (2019). Landslide susceptibility mapping using Weight of Evidence Method in Haku, Rasuwa District, Nepal. Journal of Nepal Geological Society. 58. 163-171. 10.3126/jngs.v58i0.24601

[47] Kadhim, A. I. (2019). Term Weighting for Feature Extraction on Twitter: A Comparison between BM25 and TF-IDF. 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 124 - 128). Zakho- Duhok, Iraq: IEEE Xplore

[48] Guo, Jinghuan & Mu, Yong & Xiong, Mudi & Liu, Yaqing & Gu, Jinxuan. (2019). Activity Feature Solving Based on TF-IDF for Activity Recognition in Smart Homes. Complexity. 2019. 10.1155/2019/5245373

[49] Mohammed M, Omar N (2020) Question classification based on Bloom’s taxonomy cognitive domain using modified TF-IDF and word2vec. PLOS ONE 15(3): e0230442. https://doi.org/10.1371/journal.pone.0230442.

[50] Ignacio Arroyo-Fernández, C.-F.-C. G.-M.-M. (2019). Unsupervised sentence representations as word information series: Revisiting TF-IDF. Computer Speech & Language, 107-129