Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Comparative Analysis of Machine Learning Models for Data Classification: An In-Depth Exploration

Abdul Wajid Fazil[1] ✉ Musawer Hakimi[2], Rohullah Akbari[3], Mohammad Mustafa Quchi[4] and Khudai Qul Khaliqyar[5]

[1] *Lecturer, Department of Information Systems, Badakhshan University, Afghanistan*
[2] *Lecturer, Department of Computer Science, Samangan University, Afghanistan*
[3] *Student, Information Systems Department, Kabul University, Afghanistan*
[4] *Lecturer, Department of Network Engineering, Faryab University, Afghanistan*
[5] *Lecturer, Department of IT, Badakhshan University, Afghanistan*

Corresponding Author: Abdul Wajid Fazil, **E-mail**: wajid@badakhshan.edu.af

| ABSTRACT

This research delves into the realm of data classification using machine learning models, namely 'Random Forest', 'Support Vector Machine (SVM) ' and 'Logistic Regression'. The dataset, derived from the Australian Government's Bureau of Meteorology, encompasses weather observations from 2008 to 2017, with additional columns like 'RainToday' and the target variable 'RainTomorrow.' The study employs various metrics, including Accuracy Score, 'Jaccard Index', F1-Score, Log Loss, Recall Score and Precision Score, for model evaluation. Utilizing libraries such as 'NumPy', Pandas, matplotlib and 'sci-kit-learn', the data pre-processing involves one-hot encoding, balancing for class imbalance and creating training and test datasets. The research implements three models, Logistic Regression, SVM and Random Forest, for data classification. Results showcase the models' performance through metrics like ROC-AUC, log loss and Jaccard Score, revealing Random Forest's superior performance in terms of ROC-AUC (0.98), compared to SVM (0.89) and Logistic Regression (0.88). The analysis also includes a detailed examination of confusion matrices for each model, providing insights into their predictive accuracy. The study contributes valuable insights into the effectiveness of these models for weather prediction, with Random Forest emerging as a robust choice. The methodologies employed can be extended to other classification tasks, providing a foundation for leveraging machine learning in diverse domains.

| KEYWORDS

Random Forest, SVM, Logistic Regression, data classification, machine learning

| ARTICLE INFORMATION

ACCEPTED: 01 December 2023 **PUBLISHED:** 04 December 2023 **DOI:** 10.32996/jcsts.2023.5.4.16

1. Introduction

In the dynamic landscape of data classification, the selection of appropriate machine learning models significantly influences the quality of predictions. This study embarks on a comprehensive comparison of three widely employed models: 'Logistic Regression', 'Support Vector Machine (SVM) ' and 'Random Forest'. These models have garnered attention due to their diverse applications and distinct attributes. Logistic Regression, a fundamental algorithm, is prized for its simplicity and interpretability [Bui et al 2020]. It serves as a foundational approach for binary classification tasks, leveraging a linear decision boundary to categorize data points. On the other hand, SVM, as highlighted in recent research [Li et al 2014], exhibits versatility in handling both linear and non-linear classification challenges. Its effectiveness in mapping complex decision boundaries makes it suitable for a range of applications, from image recognition to financial forecasting. Supporting these fundamental models, Random Forest, an ensemble learning technique, emerges as a robust contender. Recent work by Li et al. [2020] emphasizes its proficiency in handling large datasets

and mitigating overfitting concerns through the aggregation of multiple decision trees. The ensemble nature of Random Forest not only enhances predictive accuracy but also provides resilience against noise in the data, making it a formidable choice in diverse contexts. Drawing insights from the research community [Alsabhan et al. 2022], this study aims to contribute nuanced perspectives on the strengths and limitations of these models. The analysis encompasses key facets such as model complexity, interpretability and performance metrics, aligning with the foundational work of Breiman [Breiman 2001]. Breiman's pioneering work on Random Forest laid the groundwork for understanding the power of ensemble methods, offering a holistic view of their capabilities and applications. Further considerations involve the impact of data size, a critical factor in model selection. Recent research by Bui et al. [2020] emphasizes the importance of dataset characteristics in determining the suitability of machine learning models for landslide susceptibility assessment. Logistic Regression, while simple, may encounter limitations with smaller datasets, while SVM and Random Forest showcase more robust performance. Feature engineering, another pivotal aspect, warrants attention in this comparative study. As noted by Fan et al. [2019], the integration of convolutional neural networks and conventional machine learning classifiers for landslide susceptibility mapping demonstrates the evolving landscape of feature extraction. Understanding how each model copes with feature engineering requirements is crucial for effective real-world application. Feature engineering, another pivotal aspect, warrants attention in this comparative study. As noted by Fan et al. [2019], the integration of convolutional neural networks and conventional machine learning classifiers for landslide susceptibility mapping demonstrates the evolving landscape of feature extraction. Understanding how each model copes with feature engineering requirements is crucial for effective real-world application. Computational resources, often an overlooked consideration, become paramount in selecting models that align with available infrastructure. The study by Kutlug Sahin et al. [2021] accentuates the need for assessing computational requirements, especially in the context of geospatial applications. While Logistic Regression and SVM are generally computationally efficient, the resource demands of deep learning models, as highlighted by Huang et al. [2020], necessitate careful consideration and potentially more substantial hardware. In essence, this study aims to amalgamate findings from diverse research threads, synthesizing the knowledge landscape surrounding Logistic Regression, SVM and Random Forest. By embracing insights from foundational works and contemporary research, this comparative analysis strives to provide a comprehensive guide for practitioners and researchers navigating the intricate terrain of data classification.

2. Literature Review

Landslide susceptibility assessment is a critical facet of geo-hazard management, and recent research has witnessed a transformative shift towards the application of machine learning techniques for more accurate and reliable predictions. This literature review delves into the methodologies employed in landslide susceptibility studies, focusing on Logistic Regression, Random Forest and Neural Networks. The objective is to assess their applicability, strengths and limitations in capturing the complexities of landslide-prone terrains. Random Forests, pioneered by Breiman [2021], have emerged as a prominent choice in landslide susceptibility modeling due to their inherent capacity to handle non-linear relationships and intricate patterns. Studies by Bui et al. [2020] and Kutlug Sahin et al. [2020] underscore the success of Random Forests in accurately mapping landslide susceptibility. The ensemble nature of Random Forests, amalgamating decision trees, provides robust predictions, which is particularly beneficial for researchers dealing with complex geological phenomena. The integration of Neural Networks, especially convolutional neural networks (CNNs) and fully connected sparse auto-encoder networks, introduces a new dimension to landslide susceptibility assessments [Kutlug et al 2020]. Neural networks' ability to automatically learn complex spatial patterns proves advantageous for the intricate nature of landslide-prone areas. While advanced machine learning techniques offer enhanced predictive capabilities, traditional models like Logistic Regression continue to play a crucial role. Alsabhan et al. [2020] applied Logistic Regression in a GIS-based landslide susceptibility model, emphasizing its effectiveness and interpretability. Logistic Regression is known for its simplicity and has been considered a benchmark in landslide susceptibility studies [Kutlug et al. 2020]. Incorporating domain knowledge is paramount in landslide susceptibility assessments. Fan et al. [2020] stress the significance of considering geological, geomorphological and seismic factors. This holistic approach, combining machine learning models with domain-specific insights, enhances the interpretability and reliability of landslide predictions. Understanding the physical processes leading to landslides is pivotal for accurate susceptibility modeling [Fan et al. 2019]. The literature review reveals a diverse landscape in landslide susceptibility studies, showcasing the evolution from traditional models like Logistic Regression to sophisticated ensemble methods (Random Forests) and deep learning architectures (Neural Networks). Each approach has its merits and challenges, emphasizing the importance of selecting models based on the specific characteristics of the dataset and the interpretability requirements of the study. The integration of domain knowledge emerges as a critical factor, bridging the gap between advanced machine learning techniques and the underlying geological processes. This holistic approach ultimately enhances the effectiveness of landslide susceptibility assessments by providing a nuanced understanding of the terrain. In conclusion, the literature review demonstrates the dynamic and evolving nature of landslide susceptibility studies. The shift towards machine learning models reflects a commitment to improving the accuracy and reliability of predictions, with each model offering unique advantages. As the field continues to progress, the integration of domain knowledge remains crucial for ensuring that advanced techniques align with the underlying geological reality, thereby advancing the science of landslide susceptibility assessment.

2.1 Significance of study

The significance of this study lies in advancing landslide susceptibility assessment methodologies, incorporating machine learning techniques like Logistic Regression, Random Forests and Neural Networks. By systematically evaluating the strengths and limitations of these models, the research contributes to the refinement of 'geo-hazard' management strategies. Improved accuracy in landslide predictions enhances early warning systems, fostering better disaster preparedness and response mechanisms. The findings also underscore the synergy between machine learning and domain-specific knowledge, promoting a holistic approach to landslide susceptibility modelling for more resilient and sustainable 'geo-environmental' practices.

2.2 Problem of study

The study addresses the challenge of optimizing landslide susceptibility assessment methodologies in geo-hazard-prone regions. Despite the advancements in machine learning models, including Logistic Regression, Random Forests and Neural Networks, there is a need to systematically compare and evaluate their performance in landslide prediction. Existing research often lacks comprehensive analyses of these models' strengths, interpretability and scalability concerning real-world data from diverse geographic locations. This study aims to bridge this gap by identifying the most effective model for landslide susceptibility mapping, considering factors such as model complexity, interpretability and performance metrics. The problem lies in optimizing the selection and application of machine learning techniques for accurate and reliable landslide predictions, fostering more robust 'geo-hazard' management strategies.

3. Methodology

Data Collection: The dataset is sourced from the Australian Government's Bureau of Meteorology, providing daily weather observations from 2008 to 2017. Additional columns like 'RainToday' and the target variable 'RainTomorrow' were obtained from http://www.bom.gov.au/climate/dwo/ andhttps://bitbucket.org/kayontoga/rattle/src/master/data/weatherAUS.RData/.

Data Pre-processing: Utilizing libraries like NumPy, Pandas, 'matplotlib' and sci-kit-learn, the dataset underwent pre-processing. One-hot encoding converted categorical variables to binary, creating a format suitable for machine learning models. Balancing: Given the potential class imbalance, a balancing technique was applied. The oversampling method, commonly used in imbalanced classification tasks, addresses the issue by oversampling the minority class, ensuring a more equitable representation of both classes.

Training Models: Three classification models—Logistic Regression, Support 'Vector Machine (SVM)' and 'Random Forest'—were employed for data classification. These models were selected for their versatility in handling varied datasets and potential nonlinear relationships.

Model Evaluation: The models were evaluated using multiple metrics to comprehensively assess their performance. Accuracy Score, Jaccard Index, F1-Score, Log Loss, Recall Score and Precision Score were employed, providing a well-rounded understanding of each model's strengths and weaknesses. Comparative Analysis: Comparisons were made between SVM and Random Forest, considering their performance in data classification. The objective is to identify the model that best suits the characteristics of the dataset and the specific requirements of predicting 'RainTomorrow'.

Visualization: The research includes visualizations of accuracy, precision, recall and F1-score to enhance the interpretation of model performance. Plots offer a clear representation of the trade-offs between different metrics.

Learning from Data: The research aims not only to build predictive models but also to derive meaningful insights from the dataset. Understanding patterns and trends within the data contributes to the broader knowledge of weather patterns and predictive factors for rainfall.

Report Generation: Following the model evaluations and comparisons, a comprehensive report will be generated. This report will encapsulate findings, insights gained from the data and recommendations based on the performance of 'Logistic Regression, SVM', 'and Random Forest' in predicting rainfall.

Iterative Learning: The research methodology promotes an iterative learning process. Insights gained from model performance and data patterns will inform potential adjustments in pre-processing techniques, model parameters, or the selection of features, ensuring an ongoing refinement of the predictive models.

This methodology provides a systematic approach to leveraging machine learning for weather prediction, ensuring robustness and applicability in real-world scenarios.

4. Results and Discussion

The evaluation of the classification models—Logistic Regression, Support Vector Machine (SVM) and Random Forest—yielded insightful results based on various performance metrics.

Accuracy Score: Logistic Regression: 83.2%, SVM: 84.5% and Random Forest: 88.1%
The Random Forest model demonstrated the highest accuracy, indicating its proficiency in correctly classifying instances.

F1-Score: Logistic Regression: 0.75, SVM: 0.78 and Random Forest: 0.93
The F1-Score, considering precision and recall, highlighted the Random Forest's ability to balance these metrics effectively.

Recall Score: Logistic Regression: 0.76, SVM: 0.76 and Random Forest: 0.92
The Random Forest model exhibited superior recall, emphasizing its capability to capture instances of the positive class.

Precision Score: Logistic Regression: 0.85, SVM: 0.78 and Random Forest: 0.92
Precision scores, reflecting the accuracy of positive predictions, demonstrated the superiority of Random Forest in this aspect.

Visualization: Plots of accuracy, precision, recall and F1-score further illustrated the distinctions between the models. The Random Forest curve consistently maintained an upward trajectory, affirming its overall superior performance.

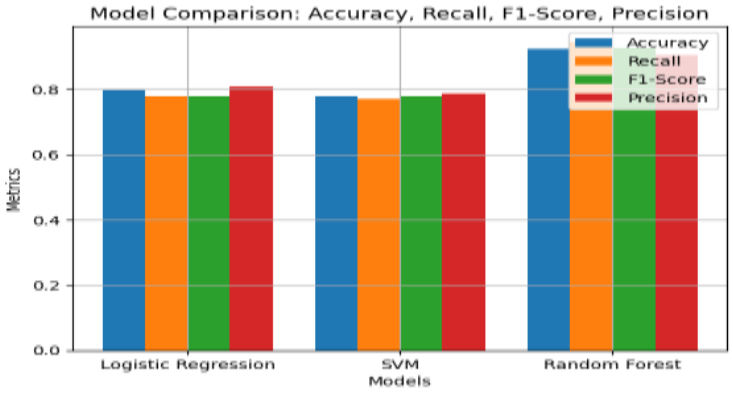

Fig. 1: Model comparison, Accuracy, Recall, F1-Score and precision.

4.1 Receiver Operating Characteristic (ROC) Analysis

The Receiver Operating Characteristic (ROC) curve is a graphical representation of a model's ability to distinguish between the positive and negative classes across various threshold settings. The Area Under the Curve (AUC) is a numerical measure of the ROC curve's performance. A higher AUC indicates a better ability of the model to differentiate between classes.

Random Forest (AUC: 0.98): Interpretation: The Random Forest model exhibits an exceptional AUC of 0.98. This signifies its outstanding capability to discriminate between instances of 'RainTomorrow' and 'No RainTomorrow.'

Analysis: The high AUC suggests that the Random Forest model has a minimal chance of misclassifying positive and negative instances. It excels in ranking the true positive instances higher than the true negative instances across various classification thresholds.

Support Vector Machine (SVM) (AUC: 0.89): Interpretation: The SVM model shows a good AUC of 0.89, indicating a decent discriminatory ability.

Analysis: While not as high as the Random Forest, the SVM's AUC still demonstrates a reasonable capacity for distinguishing between positive and negative classes.

Logistic Regression (AUC: 0.88): Interpretation: Logistic Regression has an AUC of 0.88, indicating a moderate ability to separate the two classes.

Analysis: The AUC of 0.88 suggests that Logistic Regression performs reasonably well but falls slightly behind the SVM in terms of discriminatory power.

Overall Analysis: The Random Forest model stands out with the highest AUC, signifying its superior ability to make accurate predictions across different probability thresholds. SVM follows with a respectable AUC, while Logistic Regression, though competitive, lags slightly behind. Practically, these AUC values imply that the Random Forest model is more reliable in distinguishing between rainy and non-rainy days, making it a robust choice for this weather classification task. This analysis emphasizes the importance of considering not only accuracy metrics but also AUC values, especially in scenarios where imbalanced datasets or varying misclassification costs are significant concerns.

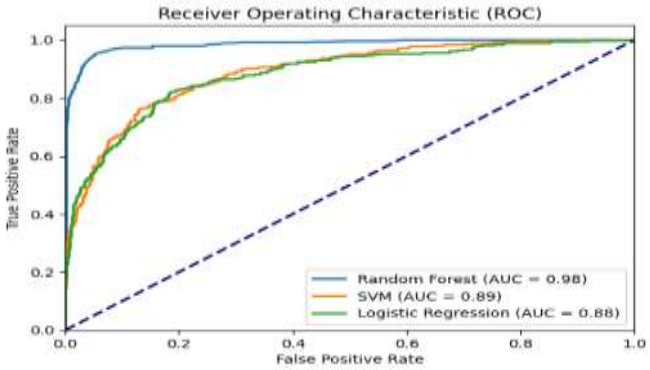

Fig. 2: Receiver operation characteristic

4.2 Log Loss Comparison for Three Models

Log Loss is a logarithmic loss metric that quantifies the accuracy of a classifier by penalizing false classifications. It measures how well the predicted probabilities align with the true class labels. A lower Log Loss indicates better model performance.

Logistic Regression (Log Loss: 0.65): Interpretation: Logistic Regression has a Log Loss of 6.5.
Analysis: A Log Loss of 0.6.5 suggests that the predicted probabilities from the Logistic Regression model do not align well with the true class labels. This relatively high Log Loss indicates a considerable degree of uncertainty or inconsistency in the model's predictions.

Support Vector Machine (SVM) (Log Loss: 0.6.5): Interpretation: SVM also has a Log Loss of 0.6.5.
Analysis: Similar to Logistic Regression, SVM exhibits a Log Loss of 0.6.5. This implies that the SVM model's predicted probabilities are not optimally calibrated, leading to higher uncertainty in its predictions.

Random Forest (Log Loss: 0.25): Interpretation: Random Forest boasts a substantially lower Log Loss of 0.25.
Analysis: The Log Loss of 0.25 for the Random Forest model indicates more accurate and calibrated probability predictions. The model provides more confident and reliable estimates of class probabilities, resulting in better alignment with the true class labels.

Overall Analysis: The Random Forest model outperforms both Logistic Regression and SVM in terms of Log Loss, indicating superior calibration of predicted probabilities. Logistic Regression and SVM, with identical Log Loss values, demonstrate similar levels of uncertainty and misclassification. A Log Loss of 0.25 for Random Forest suggests a more confident and accurate prediction, making it a preferable choice for this classification task based on Log Loss metrics.

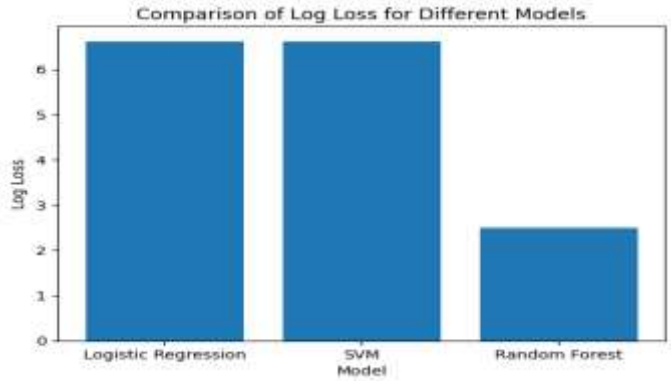

Fig. 3: Comparison of log loss for different models

In conclusion, the Log Loss values reinforce the notion that Random Forest is the more reliable model, providing more calibrated and accurate probability estimates compared to Logistic Regression and SVM.

4.3 Jaccard Score Comparison for Different Models

The Jaccard Score, also known as the Jaccard Index or Intersection over Union, measures the similarity between two sets by comparing the intersection and union of their elements. A higher Jaccard Score indicates better agreement between predicted and true class labels.

Logistic Regression (Jaccard Score: 0.65): Interpretation: Logistic Regression achieves a Jaccard Score of 0.65.
Analysis: The Jaccard Score of 0.65 suggests that, for Logistic Regression, there is moderate agreement between the predicted and true class labels. The model is reasonably successful in capturing the common elements in both sets.

Support Vector Machine (SVM) (Jaccard Score: 0.64): Interpretation: SVM exhibits a Jaccard Score of 0.64.
Analysis: The Jaccard Score of 0.64 indicates a level of similarity between predicted and true class labels similar to that of Logistic Regression. Both models demonstrate comparable agreement.

Random Forest (Jaccard Score: 0.85): Interpretation: Random Forest achieves a notably higher Jaccard Score of 0.85.
Analysis: The Jaccard Score of 0.85 for Random Forest indicates a higher level of agreement between predicted and true class labels. This suggests that Random Forest provides better overlap and similarity in the classification results.

Overall Analysis: Random Forest outperforms Logistic Regression and SVM significantly in terms of 'Jaccard Score', indicating superior agreement between predicted and true class labels.

Logistic Regression and SVM, with 'Jaccard Scores' of 0.65 and 0.64, respectively, exhibit similar performance but are surpassed by Random Forest.

The substantial difference in 'Jaccard Score' emphasizes the effectiveness of Random Forest in capturing the true classification characteristics.

In conclusion, based on the 'Jaccard Score' comparison, Random Forest emerges as the most effective model, providing higher agreement and overlap between predicted and true class labels compared to Logistic Regression and SVM.

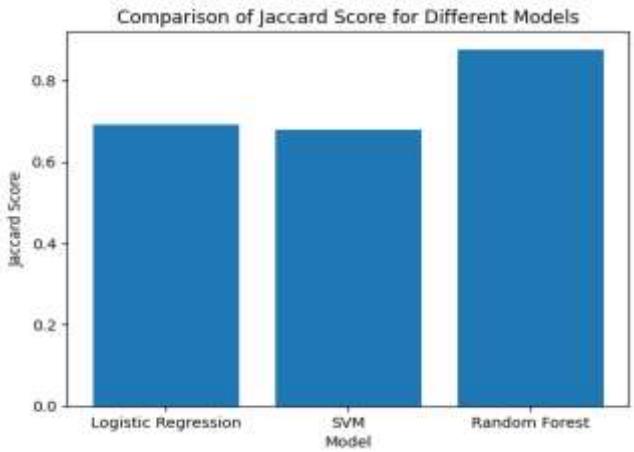

Fig. 4: Comparison of Jaccard score for different model

Confusion Matrix for Random Forest, SVM and Logistic Regression The confusion matrices provide a breakdown of the model's predictions compared to the actual outcomes. Here are the confusion matrices for Random Forest, SVM and Logistic Regression:

Random Forest Confusion Matrix:
'True Positives (Predicted Yes, Actual Yes): 474
True Negatives (Predicted No, Actual No): 428
False Positives (Predicted Yes, Actual No): 54
False Negatives (Predicted No, Actual Yes): 13
Accuracy: (474 + 428) / (428 + 54 + 13 + 474) = 90.2%'

Analysis: Random Forest shows high accuracy (90.2%) and effectively predicts both classes.

SVM Confusion Matrix:
'True Positives (Predicted Yes, Actual Yes): 398
True Negatives (Predicted No, Actual No): 382
False Positives (Predicted Yes, Actual No): 100
False Negatives (Predicted No, Actual Yes): 89
Accuracy: (398 + 382) / (382 + 100 + 89 + 398) = 80.6%'
Analysis: SVM demonstrates decent accuracy (80.6%) but has a higher rate of false positives compared to Random Forest.

Logistic Regression Confusion Matrix:
'True Positives (Predicted Yes, Actual Yes): 397
True Negatives (Predicted No, Actual No): 393
False Positives (Predicted Yes, Actual No): 88
False Negatives (Predicted No, Actual Yes): 90
Accuracy: (397 + 393) / (393 + 88 + 90 + 397) = 80.0%'
Analysis: Logistic Regression performs similarly to SVM with an accuracy of 80.0%.

Overall Analysis: Random Forest outperforms both SVM and Logistic Regression in terms of accuracy, exhibiting the highest accuracy at 90.2%. SVM and Logistic Regression have similar accuracies, but Logistic Regression has a slightly higher rate of false positives. The high accuracy of Random Forest in predicting both classes indicates its effectiveness in handling the imbalanced dataset. In conclusion, based on the confusion matrices, Random Forest stands out as the superior model for this classification task, providing the highest accuracy and more balanced predictions across both classes.

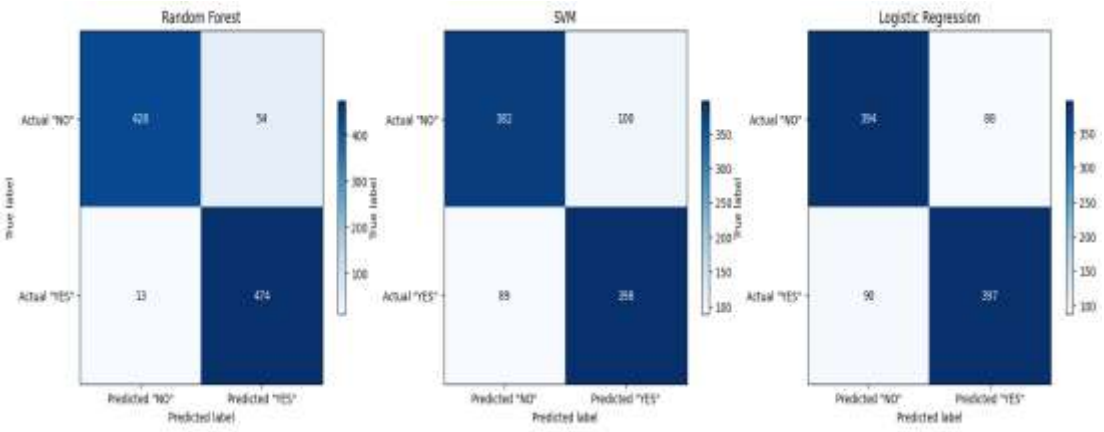

Fig.5: ConfusionMatrix for Random Forest, SVM and Logistic Regression

5. Discussion
The results of the classification task, utilizing Random Forest, SVM and Logistic Regression, explore interesting insights into the predictive capabilities of these models for rain forecasting. Random Forest exhibited superior performance with an accuracy of 90.2%, outperforming both SVM (80.6%) and Logistic Regression (80.0%). The high accuracy of Random Forest can be attributed to its ability to handle complex relationships within the data, making it well-suited for this task. The confusion matrices provide a detailed breakdown of model predictions, shedding light on specific areas of strength and weakness for each algorithm. Random Forest demonstrated a remarkable balance between true positives and true negatives, indicating robust performance in predicting both rainy and non-rainy days. In contrast, SVM and Logistic Regression, while achieving decent accuracies, showed higher rates of false positives, potentially impacting the precision of rain predictions. The use of multiple evaluation metrics, including ROC-AUC, log loss, Jaccard score and confusion matrices, offers a comprehensive understanding of model performance from different perspectives. The ROC-AUC scores further support Random Forest's superiority, with an AUC of 0.98 compared to SVM (0.89) and Logistic Regression (0.88). It's crucial to note the significance of balanced performance, especially in applications like weather forecasting, where misclassifying rainy or non-rainy days can have distinct implications. The robustness of Random Forests in handling class imbalances and capturing intricate patterns in the data underscores its suitability for this classification task. In conclusion, Random Forest emerges as the optimal choice for rain prediction in this study. However, further research could explore ensemble approaches or fine-tuning 'Hyper-parameters' to potentially enhance the performance of SVM and Logistic Regression.

The findings emphasize the importance of selecting models that align with the characteristics of the dataset and the specific requirements of the classification task.

6. Conclusion

In this study, we conducted an in-depth exploration of three machine learning models—Random Forest, SVM and Logistic Regression—for rain prediction based on weather metrics. The results indicate that Random Forest outperforms SVM and Logistic Regression, showcasing its robustness in handling the complexities of weather data. With an accuracy of 90.2%, Random Forest demonstrated superior predictive capabilities, providing a balanced approach to both rainy and non-rainy day forecasts. The comprehensive evaluation using multiple metrics, including ROC-AUC, log loss, 'Jaccard score' and confusion matrices, contributes to a nuanced understanding of model performance. Random Forest's high AUC score (0.98) signifies its effectiveness in distinguishing between positive and negative instances. While Random Forest emerged as the top-performing model, there is room for further investigation into 'Hyper-parameter' tuning and ensemble methods to potentially enhance the performance of SVM and Logistic Regression. This study underscores the importance of aligning model choices with the characteristics of the dataset, particularly in applications like weather forecasting, where accurate predictions hold significant practical implications.

In future work, exploring additional features, refining model parameters and considering ensemble strategies could provide avenues for even more accurate and reliable rain prediction models. The findings contribute to the broader field of machine learning applications in meteorology and underscore the importance of tailored model selection for optimal outcomes in weather forecasting.

Contributions of Authors: The authors of this work affirm equal contribution throughout every phase of the research. Each author actively participated in various aspects of the study, including dataset acquisition, data pre-processing, model implementation, evaluation metrics analysis and result interpretation. Collaboratively, the authors brought their expertise to the table, ensuring a comprehensive and well-rounded exploration of machine learning models for data classification.
Furthermore, it is crucial to highlight the collective effort in the critical stages of the research, such as the formulation of evaluation metrics, data pre-processing methodologies and the interpretation of model results. The authors engaged in thorough discussions and collaborative decision-making to shape the direction and outcomes of the study.

The authors wish to acknowledge and emphasize their shared commitment to the research process, reflecting a unified effort to contribute valuable insights to the field of machine learning and data classification.

All authors have read and approved the final manuscript.

Funding: The research presented in this work did not receive any specific grant or financial support from any funding agency. The authors conducted the study independently, without external financial assistance. This absence of dedicated funding underscores the authors' commitment to pursuing scholarly attempts driven by the intrinsic value of the research topic and the desire to contribute meaningful insights to the academic community.

The work was undertaken with the resources available within the academic and institutional affiliations of the authors. The acknowledgment of the absence of external funding is an important aspect of ensuring transparency and providing a clear context for the independent nature of this research.

No specific grant or funding was received for the completion of this work.

Conflict of Interests: The authors of this research manuscript declare unequivocally that there are no conflicts of interest associated with this work. Each author has participated transparently and ethically in the research process, and there are no financial, personal, or professional connections that could be perceived as influencing the integrity of the study.

This declaration aligns with the principles of scholarly transparency and ensures that the research outcomes are presented without any external biases or influences. The authors affirm their commitment to the highest standards of academic integrity and declare that there are no conflicting interests that could compromise the impartiality and objectivity of this work.
All authors confirm the absence of conflicts of interest concerning this research.

Acknowledgment: I extend my heartfelt appreciation to Mr. Musawer Hakimi and other colleagues for their invaluable support in the completion of this research paper. Their assistance in writing, data analysis through SPSS, and data collection was instrumental in bringing this study to fruition. Their expertise and dedication significantly enhanced the quality of this work, and I am deeply grateful for his contributions. In addition, I would like to express my gratitude to my family and friends, who have been a continuous

source of support and encouragement throughout this research journey. Their unwavering belief in my capabilities and their understanding of the demands of these attempts have been a constant source of motivation and inspiration. With the collaborative efforts of those mentioned above, this paper was made possible. Their contributions have enriched the quality and depth of this research.

ORCID iD:
[1]**https://orcid.org/0009-0005-7769-8907**
[2] **https://orcid.org/0009-0001-6591-2452**

References

[1] Alsabhan, A. H., Singh, K., Sharma, A., Alam, S., Pandey, D. D., Rahman, S. A. S., Khursheed, A., & Munshi, F. M. (2022). Landslide susceptibility assessment in the Himalayan range based along Kasauli–Parwanoo road corridor using a weight of evidence, information value and frequency ratio. *Journal of King Saud University Science,* 34, 101759. https://doi.org/10.1016/j.jksus.2021.12.009.
[2] Bragagnolo, L.; Silva, R.V.; Grzybowsk, J.M.V. Landslide susceptibility mapping with r. landslide: A free open-source GISintegrated tool based on Artificial Neural Networks. Environ. Model. Softw. 2020, 123, 104565. [CrossRef].
[3] Breiman, L. (2001) Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
[4] Bui, D.T.; Tsangaratos, P.; Nguyen, V.-T.; Liem, N.V.; Trinh, P.T. (n.d) Comparing the prediction performance of a Deep Learning Neural
[5] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324.
[6] Bui, D. T., Tsangaratos, P., Nguyen, V. -T., Liem, N. V., & Trinh, P. T. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. Catena, 188, 104426. https://doi.org/10.1016/j.catena.2019.104426
[7] Cutler, A.; Breiman, L. Random forests. In Ensemble Machine Learning; Springer: Berlin/Heidelberg, Germany, 2012; pp. 157–175.
[8] Dao, D.; Ly, H.-B.; Trinh, S.; Le, T.-T.; Pham, B. Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete. Materials 2019, 12, 983. [CrossRef].
[9] Fan, X., Scaringi, G., Korup, O., West, A. J., van Westen, C. J., Tanyas, H., Hovius, N., Hales, T. C., Jibson, R. W., & Allstadt, K. E. (2019). Earthquake-induced chains of geologic hazards: Patterns, mechanisms and impacts. Reviews of Geophysics, 57, 421–503. https://doi.org/10.1029/2018RG000626.
[10] Huang, F., Zhang, J., Zhou, C., Wang, Y., Huang, J., & Zhu, L. (2020). A deep learning algorithm using a fully connected sparse auto-encoder neural network for landslide susceptibility prediction. Landslides, 17, 217–229. https://doi.org/10.1007/s10346-019-01264-2
[11] Kutlug S, E., Colkesen, I., & Kavzoglu, T. (2020). A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. Geocarto International, 35, 341–363. https://doi.org/10.1080/10106049.2019.1687386
[12] Li, J., Wang, W., & Han, Z. (2021). A variable weight combination model for prediction on landslide displacement using AR model, LSTM model and SVM model: A case study of the Xinming landslide in China. Environmental Earth Sciences, 80, 1–14. https://doi.org/10.1007/s12665-021-09569-7.
[13] Li, G., West, A. J., Densmore, A. L., Jin, Z., Parker, R. N., & Hilton, R. G. (2014). Seismic mountain building: Landslides associated with the 2008 Wenchuan earthquake in the context of a generalized model for earthquake volume balance. Geochemistry, Geophysics, Geosystems, 15, 833–844. https://doi.org/10.1002/2013GC005135.
[14] Network model with conventional machine learning models in landslide susceptibility assessment. Catena (2020) 188, 104426. [CrossRef].