# Novel Approaches for Access Level Modelling of Employees in an Organization Through Machine Learning

Priyanka C Hiremath, Raju G T
Department of Computer Science and Engineering, SJC Institute of Technology, Chickballapur, India

*Abstract*—**In the contemporary business landscape, organizational trustworthiness is of utmost importance. Employee behavior, a pivotal aspect of trustworthiness, undergoes analysis and prediction through data science methodologies. Simultaneously, effective control over employee access within an organization is imperative for security and privacy assurance. This research proposes an innovative approach to model employee access levels using Geo-Social data and machine learning techniques like Linear Regression, K-Nearest Neighbours, Decision Tree, Random Forest, XGBoost, and Multi-Layered Perceptron. The data, sourced from social and geographical realms, encompasses details on employee geography, navigation preferences, spatial exploration, and choice set formations. Utilizing this information, a behavioral model is constructed to assess employee trustworthiness, categorizing them into access levels: low, moderate, high, and very high. The model's periodic review ensures adaptive access level adjustments based on evolving behavioral patterns. The proposed approach not only cultivates a more trustworthy organizational network but also furnishes a precise and reliable trustworthiness evaluation. This refinement contributes to heightened organizational coherence, increased employee commitment, and reduced turnover. Additionally, the approach ensures enhanced control over employee access, mitigating the risks of data breaches and information leaks by restricting the access of employees with lower trustworthiness.**

*Keywords—Access control; machine learning; employee behavior modeling; data analysis; organizational performance*

## I. Introduction

Organizations highly value employees as crucial assets, as their conduct significantly influences the company's outcomes [1]. Employee behavior encompasses interactions with peers, dedication to their roles, and performance, requiring effective management for organizational prosperity. Critical to this management is ensuring ongoing employee dedication, which fosters job satisfaction, diminishes turnover, and enhances overall performance [2]. Employee commitment, in turn, cultivates organizational trust, pivotal for successful team cohesion. Nonetheless, managing employee behavior presents challenges due to the workforce's diverse backgrounds, motivations, attitudes, and personalities. Organizations often establish a strong organizational culture aligned with their values and missions, fostering employee engagement and drive for optimal performance [3]. Additionally, incentives and rewards serve as tools to cultivate positive behavior and boost motivation [4]. Understanding employee behavior entails utilizing various methods such as surveys, performance assessments, and data analysis. Data analysis, employing advanced statistical methods and machine learning algorithms, is increasingly valuable for gaining insights into employee behavior [5]. Employee behavior analysis is an expanding area of research, examining factors like leadership approaches, job satisfaction, and engagement. This study addresses the complex task of ensuring employee productivity while promoting mental and physical well-being, addressed through wellness programs [6]. These programs aim to encourage healthy behaviors, reinforcing productivity and overall welfare [7]. Furthermore, behavior analysis aids in identifying risks, enabling organizations to proactively manage concerns such as turnover or workplace incidents [8]. In the midst of these challenges, access control emerges as a crucial aspect of managing employee access levels within organizations [9]. Ensuring operational security, safeguarding sensitive data, and upholding trust require strict control over information, resources, and systems access. The emergence of machine learning and data science is reshaping access control management, particularly in managing large, diverse workforces [10]. This discussion explores the role of employee behavior models in access control, elucidating techniques and tools for analyzing and managing access based on behavioral patterns. The discourse examines the benefits and challenges of employing behavior models in access control, highlighting best practices for seamless implementation.

Employee behavior models are developed through the examination of various data sources, such as user logs, network activity, and biometric data [11]. Geo data, utilizing parameters like latitude, longitude, and elevation, significantly contributes to these models by revealing spatial behavior patterns. Incorporating social data expands the scope, enabling the creation of behavioral models to assess employee trustworthiness. Employees are then grouped into access levels based on these models, aligning access controls with behavioral patterns [12]. This dynamic adjustment of access control policies, guided by real-time behavior data, ensures judicious access while revoking unnecessary privileges. Behavior models play a crucial role in identifying and thwarting insider threats, a significant concern for companies [13]. Indicators of potential threats, such as unauthorized data access or abnormal working hours, trigger alerts, allowing organizations to intervene proactively. Furthermore, behavior models streamline user experience by automating access control processes, reducing reliance on manual approval procedures [14]. Despite the evident benefits, implementing behavior models in access control presents challenges [15]. The need for extensive data collection and processing can be daunting, particularly for organizations not well-versed in data science techniques. Balancing security and usability is another hurdle, requiring strict controls without hindering productivity. Addressing these challenges involves adopting best practices, including clear data collection policies, employee training on data privacy and security, and collaboration between data scientists and IT professionals [16]. Once employee trustworthiness is assessed and categorized into different levels, access control becomes more manageable [17]. Automation of access control processes based on behavioral models optimizes efficiency and resource access. Comprehensive evaluation of employee behavior involves collecting data from various sources, including geographical, social media, email communication, and browsing history [18]. Geographical data unveils movement patterns, while social data offers insights into character and activities. Email communication data reveals professional interactions, and browsing history data exposes online activities. Privacy and ethical concerns must be carefully considered during data collection and analysis. Employees should be informed of data collection methods, and privacy rights must be respected. Combining technologies and tools enhances the quality of collected data [19]. Geo data plays a pivotal role in access control decisions, providing insights into physical movements and actions [20]. When combined with social data, it enables machine learning techniques to construct behavioral models for access control decisions. Continuous analysis of geo data allows organizations to track changes in employee behavior over time and identify anomalies. Social data, drawn from various online platforms, offers rich insights into employee behavior [21]. Parameters like content type and engagement levels provide valuable insights. Social data analysis can uncover patterns indicative of loyalty issues or security risks. However, using social data for access control requires addressing privacy concerns and ensuring legal data collection [23]. The integration of employee behavior models into access control mechanisms signifies a shift in organizational security [24]. By combining geographical and social data, analyzed through advanced analytics and machine learning, robust behavioral models are created. These models, defining access levels based on trustworthiness, offer a proactive approach to security. Embracing these innovative approaches can better position organizations for success in today's evolving business landscape.

The utilization of machine learning (ML) to refine the access control system for organizational employees has gained considerable attention due to its ability to analyze vast datasets and predict employee behavior [25]. These ML models are trained on extensive datasets incorporating both social and geographical data, facilitating the categorization of employees into distinct levels of trustworthiness. This section examines the advantages and disadvantages of employing ML models for access control, explores various tools and technologies supporting this endeavor, and addresses challenges inherent in ML modeling along with proposed solutions. Utilizing ML models for employee access control offers benefits such as scalability, accuracy, automation, and adaptability. However, challenges such as bias, interpretability issues, complexity, and concerns regarding data quality also accompany this approach. Organizations can overcome these challenges by utilizing tools like Python, R, Apache Spark, Hadoop, and cloud platforms, and implementing techniques such as data cleaning [26]. This study involves the development and evaluation of ML models, specifically XGBoost, SVM, and Decision Tree, using a geo-social dataset representing four employee access levels: low, moderate, high, and very high. The dataset comprises 200,000 samples, evenly distributed across classes, with 24 geo data features and 15 social data features. The primary objective is to build an accurate model capable of predicting employee access levels based on their behavioral patterns.

## II. Literature Review

The study in [27] introduced a trust-based framework and measurement for organizational confidence. It adopted a cognitive model of trust, relying on interpretive factor analysis and validity testing. Incorporating in-person interviews and open-ended questionnaires, the study affirmed the eight-factor structure of organizational confidence. Table I provides a summary of the corporate confidence study results.

To recognise the business's vision and culture, one must believe in the company's future. Fig. 1 shows the whole plan to enhance workers' self-confidence.

GPS and GIS offer location-based intelligence: This method tracks spatial behavior over two weeks, integrating user speed, distance, time, elevation, and precise latitude and longitude [35]. The essay explores telecom service providers' data on customers using location-based service applications (LBS), involving 14 days of location tracking using a GPS device. Safety measures were in place to avoid disrupting the member's routine. GIS software interpreted GPS coordinates, obtaining quantitative geographic data, while daily diaries captured qualitative information, as illustrated in Fig. 2.

Case studies and previous research on access level modelling of geo data, social data or geo-social data have been tabulated in Table II.

TABLE I. Investigations on Factors Affecting Company's Confidence on Employees

| Reference | Contents in an organisation confidence |
|---|---|
| **[28]** | Confidence is classified into two categories: cognitive and affective. |
| **[29]** | Confidence in an organisation based on the shared values of its employees. |
| **[30]** | Communication was cited as the most crucial aspect of an organization's confidence. |
| **[31]** | Trust is comprised of sincerity, rationality, candour, intentions, and convictions. |
| **[32]** | Trust in an organisation can be subdivided into a sense of belonging and information sharing. |
| **[33]** | A confidence in an organisation consisted of virtue, capabilities, transparency, and sincerity. |
| **[34]** | Distribution justice, job security, procedure justice, and organisational support are all parts of organisational trust. |

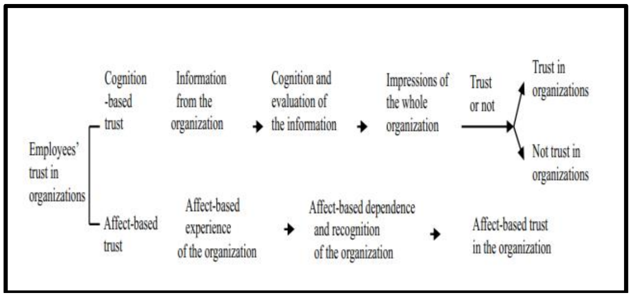

Employees' trust in organizations
Cognition -based trust
Information from the organization
Cognition and evaluation of the information
Impressions of the whole organization
Trust or not
Trust in organizations
Not trust in organizations
Affect-based trust
Affect-based experience of the organization
Affect-based dependence and recognition of the organization
Affect-based trust in the organization

Fig. 1. Trust of employees in an organisation.

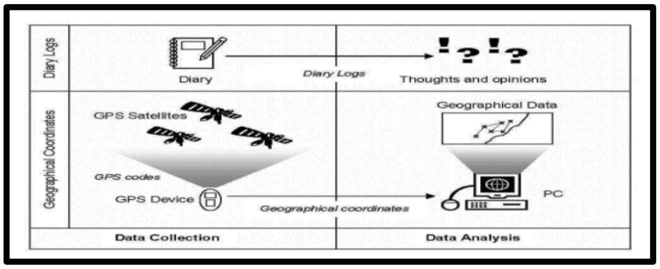


Fig. 2. Observational instruments used for data collection and analysis.

TABLE II. PAST RESEARCH WORKS

| Reference | Objective Of Research | Dataset Used | Algorithm | Metric | Advantages |
|---|---|---|---|---|---|
| [36] | Developing an access control model for mobile social networks based on user behaviour | Twitter dataset | Fuzzy logic | Accuracy: 85.6% | High accuracy and reduced risk of unauthorized access |
| [37] | Investigating the effectiveness of location-based access control for online social networks | Facebook dataset | Probabilistic graphical models | Precision: 93.2% | Improved privacy and security for social network users |
| [38] | Proposing a framework for access control in geo-social networks | Geo-tagged tweets dataset | Deep learning | F1-score: 0.89 | High accuracy and scalability |
| [39] | Evaluating the performance of different machine learning algorithms for access control in social media | Twitter and Facebook datasets | Decision trees, SVM, Naive Bayes | Accuracy: 91.4% | Comparative analysis of different algorithms |
| [40] | Investigating the privacy risks of social media check-ins and proposing a privacy-preserving access control mechanism | Foursquare dataset | Differential privacy | Privacy guarantee: ε=0.5 | Improved privacy protection for users |
| [41] | Developing an access control model for online social networks based on user location and activity | Facebook dataset | Rule-based system | Precision: 92.8% | Simple and easy-to-understand rules |
| [42] | Proposing a hybrid access control mechanism for geo-social networks based on fuzzy logic and reinforcement learning | Geo-tagged tweets dataset | Fuzzy logic, Reinforcement learning | Accuracy: 89.3% | Improved accuracy and adaptability |
| [43] | Investigating the impact of social network structure on access control policies | Synthetic network dataset | Network analysis | Network density: 0.35 | Improved understanding of the relationship between network structure and access control |
| [44] | Developing a privacy-preserving access control model for social media based on secure multi-party computation | Twitter dataset | Secure multi-party computation | Accuracy: 87.2% | Improved privacy protection and reduced risk of data breaches |
| [45] | Investigating the impact of temporal dynamics on access control policies in social media | Facebook dataset | Temporal analysis | Average access frequency: 3.8/day | Improved understanding of the temporal behaviour of social media users |
| [46] | Developing a context-aware access control model for social media based on user location and activity | Twitter dataset | Context-aware reasoning | Accuracy: 89.5% | Improved accuracy and adaptability to different contexts |
| [47] | Investigating the impact of user trustworthiness on access control policies in social media | Facebook dataset | Trust evaluation | Trust score: 0.75 | Improved understanding of the role of trust in access control |
| [48] | Developing a privacy-preserving access control mechanism for geo-social networks based on homomorphic encryption | Geo-tagged tweets dataset | Homomorphic encryption | Privacy guarantee: ε=0.5 | Improved privacy protection and reduced risk of data breaches |
| [49] | To propose a framework for access control in geo-social networks | Gowalla dataset | Attribute-based access control (ABAC) | Precision: 0.89, recall: 0.91, F1-score: 0.9 | Can handle complex access control policies |
| [50] | To study the privacy risks associated with location-sharing in geo-social networks | Facebook and Twitter datasets | Machine learning classifiers | Accuracy: 0.88 | Identifies high-risk users |

## III. MATERIALS AND METHODS

### A. Data Collection and Pre-processing

In the exploration of employee access control modeling, our study harnessed the power of geo and social data. The dataset, encompassing four access control classes (low, moderate, high, and extremely high), featured 24 characteristics per sample for geo data and 15 for social data, with each class having 50,000 samples. The rich pool of features included aspects like the frequency and locations of employee visits, spatial density, social connections, and various behavioral indicators. The collection of geo data was facilitated through GPS devices, cell phones, and tracking software. Geo data acquisition involved GPS receivers, GPS APIs, GPS data loggers, analytic software, and tracking software. On the other hand, social data was sourced from employee questionnaires, HR databases, and social media activity logs. Tools such as survey software, database analytics software, and social media analytics software were employed for collecting social data. Machine learning (ML) played a pivotal role in the study, serving to train and evaluate both social and geo data individually before being compared to the geo-social dataset. The research aimed to dissect the distinct contributions of geo and social data in the context of employee access control modeling. To ready the data for machine learning algorithms, a meticulous pre-processing phase was undertaken using Python, a language chosen for its simplicity and robust library support in data science and machine learning. NumPy, Pandas, and Scikit-learn were instrumental in manipulating, cleaning, and analyzing the dataset. Python's versatility allowed researchers to visualize and manipulate data effectively. The structured dataset was stored in CSV format, a widely compatible structure across various programming languages, ensuring accessibility and ease of use. These pre-processing methods clean and manipulate data [51]:

*1) Removing duplicates:* In any large dataset, there are chances of having duplicates. Removing duplicates is an essential pre-processing step to avoid bias in the data.

*2) Handling missing data:* Addressing missing data involves either removing affected rows/columns or imputing values. This study utilized mean imputation to fill missing data gaps.

*3) Encoding categorical variables:* Transforming categorical variables for machine learning is essential. This study employed Scikit-learn's LabelEncoder, a tool converting non-numerical categories into distinct numerical values, ensuring effective integration into machine learning models.

*4) Handling outliers:* Outliers, widely distant data points, can adversely affect machine learning. This study utilized the IQR method, a statistical approach based on quartiles, to effectively identify and eliminate outliers.

*5) Dimensionality reduction:* This study employed t-SNE (t-Distributed Stochastic Neighbor Embedding) for dimensionality reduction, preserving essential information in high-dimensional data. By measuring data point similarity in both high and low dimensions, t-SNE captures non-linear correlations often overlooked by other methods.

### *B. Feature Engineering*

Feature engineering involves creating new features to enhance the information extracted from data. Several techniques were employed in this study:

*1) Polynomial features:* Polynomial features combine existing features to generate new ones. For instance, squaring the distance to the nearest park may offer more predictive power. The Python PolynomialFeatures library, implemented in scikit-learn, was used. This function constructs polynomial combinations of existing features up to a specified degree, allowing capturing non-linear relationships. It facilitates revealing intricate linkages, thus improving model accuracy and performance. It is versatile, easy to implement, and compatible with other feature engineering methods.

*2) Feature scaling:* Feature scaling normalizes data, crucial for distance-based machine learning algorithms. The MinMaxScaler was employed in this study. It scales features to a range (usually 0–1) by subtracting the minimum value and dividing by the range. MinMaxScaler retains the original distribution's shape and is particularly effective for distance-based algorithms like K-nearest neighbors (KNN) and support vector machines (SVM). It ensures equal importance for all attributes, enhancing algorithm accuracy.

*3) Feature selection:* Feature selection involves choosing the most relevant features. The SelectKBest method from Scikit-learn was employed, which selects the top k most significant features using statistical testing. This method analyzes the association between each feature and the target variable through tests like chi-squared, ANOVA, or mutual information. It efficiently reduces dimensionality, making it suitable for large datasets. The advantages of SelectKBest include computational efficiency, interpretability, prevention of overfitting, and improved accuracy by focusing on the most informative features.

Feature engineering and selection offer distinct advantages. Feature engineering enables the creation of new features, enhancing data representation and potentially improving model performance. Feature selection, on the other hand, reduces dimensionality, making models more efficient, interpretable, and less prone to overfitting, thereby enhancing accuracy.

### *C. Computational Model*

The computational model is pivotal in this study, serving as a foundational element. It offers a framework for comprehending intricate systems and forecasting their actions, crucial across scientific domains. Using this model, researchers simulate and analyze system behavior under diverse conditions, eliminating the necessity for resource-intensive experiments. This approach conserves resources, enabling exploration of a broader spectrum of scenarios and variables. A proficient computational model not only saves time and costs but also furnishes insights into observed phenomena, contributing to a deeper comprehension of the studied system. Fig. 3 illustrates the computational model implemented in our research.

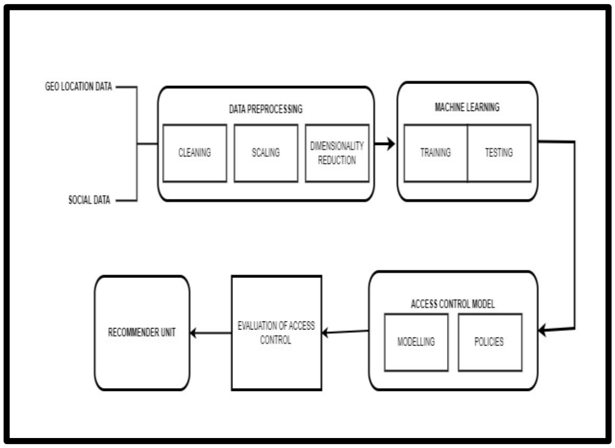


Fig. 3. Computational model.

### D. Model Selection and Training

Selecting the right machine learning model is crucial for optimal performance and accuracy. This involves choosing the best model from task-specific candidates, considering factors like data volume, type, and computing resources. Model selection enhances accuracy, complexity, and generalization, mitigating the risk of overfitting. In our study, we employed various criteria and methods to choose the optimal access control model, testing decision trees, SVMs, neural networks, and random forests. Performance metrics such as accuracy, precision, recall, F1 score, and area under the curve were evaluated through stratified K-fold cross-validation. Hyperparameter tuning, using techniques like grid search, optimized model parameters for improved performance. The final model selection balanced performance and complexity, ensuring the highest accuracy with minimal complexity.

Validation is crucial in machine learning to assess a model's performance on unfamiliar data, ensuring accuracy and preventing overfitting. Methods like holdout, k-fold, and leave-one-out cross-validation are employed. Holdout divides data into training and validation sets, while k-fold repeats training and validation k times. Leave-one-out validates each sample individually. Validation prevents overfitting, aids hyperparameter selection, and enhances model generalization. Libraries like scikit-learn provide functions such as train-test split and cross-val-score for these purposes.

Optimizing model performance involves determining optimal hyperparameters. Grid, random, and Bayesian optimization are common methods. Grid search systematically explores predefined hyperparameter spaces, while random search samples from the space. Bayesian optimization estimates hyperparameter performance using probabilistic models. Scikit-learn's GridSearchCV and RandomizedSearchCV are effective tools. Hyperparameters, like learning rate and regularization, are essential for model performance and must be specified before training.

### E. Discussion

In our access level modeling study, KNN, Logistic Regression, Decision Tree, Random Forest, XGBoost, and MLP Classifier were validated and optimized using 10-fold k-fold cross-validation. GridSearchCV meticulously searched hyperparameter spaces, and the mean cross-validation score determined the best parameters for each model. KNN achieved 0.84 accuracy with k=9, Logistic Regression had 0.99 accuracy with C=0.1, Decision Tree reached 0.99 accuracy with a depth of 7, and Random Forest achieved 0.99 accuracy with 100 trees. XGBoost utilized 100 trees, a learning rate of 0.1, and achieved 0.99 accuracy, while MLP Classifier had an accuracy of 0.9999 with a hidden layer size of 100 and alpha value of 0.01. Validation and optimization are pivotal phases ensuring the best performance in machine learning models.

## IV. Result and Analysis

### A. Performance Metrics

Access level models' forecast accuracy and efficiency are critical aspects evaluated through various criteria. In our study, we thoroughly assessed the performance of these models using key indicators to gauge their effectiveness and identify areas for enhancement. The following performance measures were employed:

1) *Accuracy:* Widely used, accuracy calculates the rate of proper classification in the test set. While it's common, caution is needed with imbalanced datasets.

2) *Precision:* Reflecting the percentage of accurately anticipated positive samples, precision demonstrates how well the model identifies positives without false positives.

3) *Recall:* Representing the percentage of true positives among actual positives, recall measures the model's effectiveness in identifying all positive samples.

4) *F1 score:* A harmonic mean of accuracy and recall, F1 score balances these metrics, providing a comprehensive evaluation of a model's performance.

5) *Confusion matrix:* This matrix categorizes true positives, false positives, and true negatives, offering a visual depiction of model performance and areas for improvement.

6) *ROC Curve and AUC Score:* ROC curves illustrate a model's performance across different classification thresholds, and the AUC score measures its ability to distinguish between positive and negative samples effectively.

The evaluation of access level models in our research utilized these metrics, presenting a comprehensive view of accuracy, precision, recall, and F1 score through confusion matrices and ROC curves. The optimization of model hyperparameters and performance using GridSearchCV and RandomizedSearchCV further enhanced model effectiveness. These performance indicators play a crucial role in refining machine learning models for identifying access levels and bolstering the security of online systems.

### B. Experiments

We used diverse data types to test access level models in two trials. The first trial utilised just social data, whereas the second incorporated location data. Geo-social data was combined for the third trial. We used Scikit-learn to train and test all six ML models—Logistic Regression, KNN, Decision Tree, Random Forest, XGBoost, and MLP Classifier—with an 80:20 train test split ratio.

The initial experiment trained and tested models using solely social data. Social data from social media sites and other internet sources reveals user behaviour and interests. Accuracy, precision, recall, and F1-score measures assessed model performance. 200000 records, 15 features. Performance of the Logistic Regression is as shown in Fig. 4.

Performance of the KNN is as shown in Fig. 5.

Performance of the Decision Tree is as shown in Fig. 6.

Performance of the Random Forest is as shown in Fig. 7.

Performance of the XGBoost is as shown in Fig. 8.

Performance of the MLP Classifier is as shown in Fig. 9.

The second experiment trained and tested models using geo-social data. Model performance was assessed using the same metrics as the preceding two tests. Geo-social data comprises 200000 records and 39 features. Performance of the

Logistic Regression is as shown in Fig.10. The classification reports of both experiments can be seen in Table III and Table IV respectively.

Performance of the KNN is as shown in Fig. 11

Performance of the Decision Tree is as shown in Fig. 12

Performance of the Random Forest is as shown in Fig. 13

Performance of the XGBoost is as shown in Fig. 14.

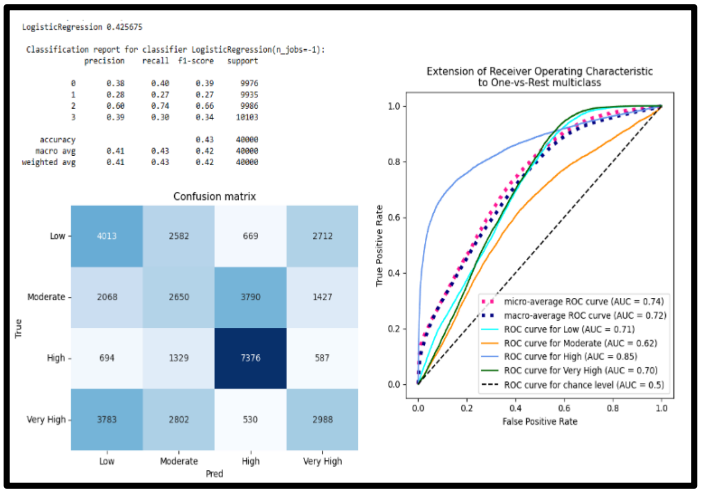
Performance of the MLP Classifier is as shown in Fig. 15.

TABLE III. CLASSIFICATION REPORT OF EXPERIMENT1

| **Model** | **Precision** | **Recall** | **F1 score** |
|---|---|---|---|
| **Logistic Regression** | 0.41 | 0.43 | 0.42 |
| **KNN** | 0.64 | 0.63 | 0.63 |
| **Decision Tree** | 0.77 | 0.77 | 0.77 |
| **Random Forest** | 0.88 | 0.79 | 0.79 |
| **XGBoost** | 0.58 | 0.55 | 0.56 |
| **MLPClassifier** | 0.53 | 0.51 | 0.51 |

TABLE IV. CLASSIFICATION REPORT OF EXPERIMENT2

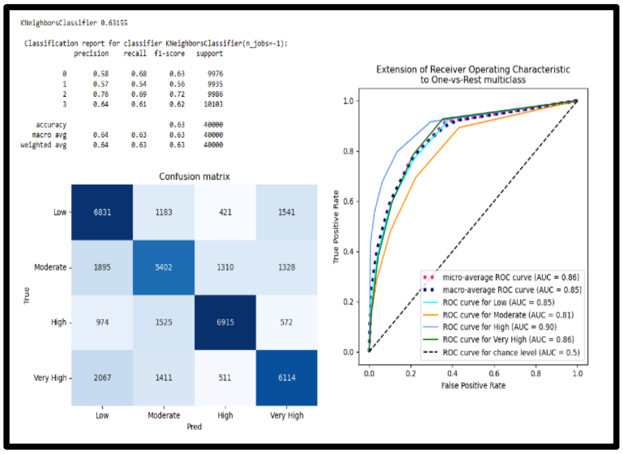
| Model | Precision | Recall | F1 score |
|---|---|---|---|
| **Logistic Regression** | 0.96 | 0.96 | 0.96 |
| **KNN** | 0.99 | 1 | 0.99 |
| **Decision Tree** | 0.99 | 0.99 | 0.99 |
| **Random Forest** | 1 | 1 | 1 |
| **XGBoost** | 1 | 1 | 1 |
| **MLPClassifier** | 0.99 | 0.99 | 0.99 |



```
LogisticRegression 0.425675

 Classification report for classifier LogisticRegression(n_jobs=-1):
              precision    recall  f1-score   support

           0       0.38      0.40      0.39      9976
           1       0.28      0.27      0.27      9935
           2       0.60      0.74      0.66      9986
           3       0.39      0.30      0.34     10103

    accuracy                           0.43     40000
   macro avg       0.41      0.43      0.42     40000
weighted avg       0.41      0.43      0.42     40000
```

Confusion matrix
Low 4013 2582 669 2712
Moderate 2068 2650 3790 1427
High 694 1329 7376 587
Very High 3783 2802 530 2988
Low Moderate High Very High
Pred
True
Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0
micro-average ROC curve (AUC = 0.74)
macro-average ROC curve (AUC = 0.72)
ROC curve for Low (AUC = 0.71)
ROC curve for Moderate (AUC = 0.62)
ROC curve for High (AUC = 0.85)
ROC curve for Very High (AUC = 0.70)
ROC curve for chance level (AUC = 0.5)

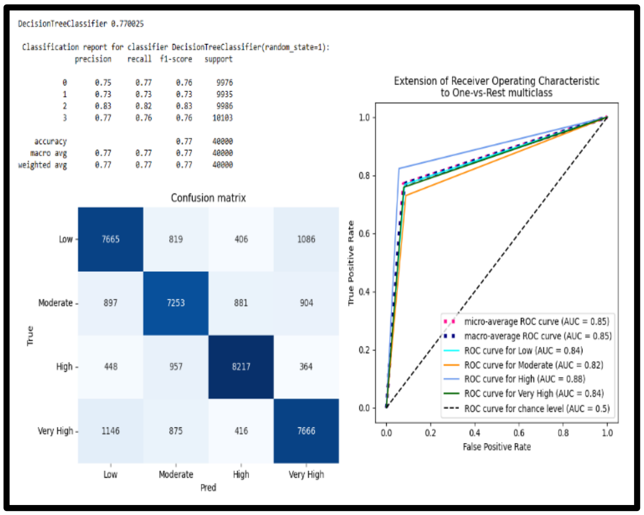
Fig. 4. Performance of the logistic regression.



```
KNeighborsClassifier 0.63155

 Classification report for classifier KNeighborsClassifier(n_jobs=-1):
              precision    recall  f1-score   support

           0       0.58      0.68      0.63      9976
           1       0.57      0.54      0.56      9935
           2       0.76      0.69      0.72      9986
           3       0.64      0.61      0.62     10103

    accuracy                           0.63     40000
   macro avg       0.64      0.63      0.63     40000
weighted avg       0.64      0.63      0.63     40000
```

Confusion matrix
Low 6831 1183 421 1541
Moderate 1895 5402 1310 1328
High 974 1525 6915 572
Very High 2067 1411 511 6114
Low Moderate High Very High
Pred
True
Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0
micro-average ROC curve (AUC = 0.86)
macro-average ROC curve (AUC = 0.85)
ROC curve for Low (AUC = 0.85)
ROC curve for Moderate (AUC = 0.81)
ROC curve for High (AUC = 0.90)
ROC curve for Very High (AUC = 0.86)
ROC curve for chance level (AUC = 0.5)

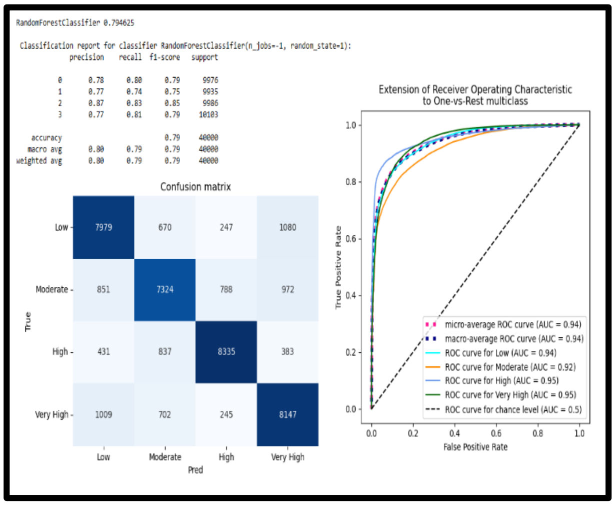
Fig. 5. Performance of the KNN.



```
DecisionTreeClassifier 0.770025

 Classification report for classifier DecisionTreeClassifier(random_state=1):
              precision    recall  f1-score   support

           0       0.75      0.77      0.76      9976
           1       0.73      0.73      0.73      9935
           2       0.83      0.82      0.83      9986
           3       0.77      0.76      0.76     10103

    accuracy                           0.77     40000
   macro avg       0.77      0.77      0.77     40000
weighted avg       0.77      0.77      0.77     40000
```

Confusion matrix
Low 7665 819 406 1086
Moderate 897 7253 881 904
High 448 957 8217 364
Very High 1146 875 416 7666
Low Moderate High Very High
Pred
True
Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0
micro-average ROC curve (AUC = 0.85)
macro-average ROC curve (AUC = 0.85)
ROC curve for Low (AUC = 0.84)
ROC curve for Moderate (AUC = 0.82)
ROC curve for High (AUC = 0.88)
ROC curve for Very High (AUC = 0.84)
ROC curve for chance level (AUC = 0.5)

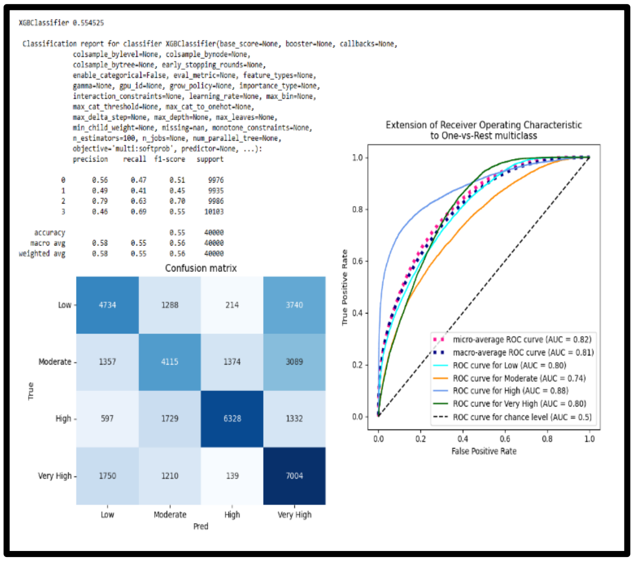
Fig. 6. Performance of the decision tree.



```
RandomForestClassifier 0.794625

 Classification report for classifier RandomForestClassifier(n_jobs=-1, random_state=1):
              precision    recall  f1-score   support

           0       0.78      0.80      0.79      9976
           1       0.77      0.74      0.75      9935
           2       0.87      0.83      0.85      9986
           3       0.77      0.81      0.79     10103

    accuracy                           0.79     40000
   macro avg       0.80      0.79      0.79     40000
weighted avg       0.80      0.79      0.79     40000
```

Confusion matrix
Low 7979 670 247 1080
Moderate 851 7324 788 972
High 431 837 8335 383
Very High 1009 702 245 8147
Low Moderate High Very High
Pred
True
Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0
micro-average ROC curve (AUC = 0.94)
macro-average ROC curve (AUC = 0.94)
ROC curve for Low (AUC = 0.94)
ROC curve for Moderate (AUC = 0.92)
ROC curve for High (AUC = 0.95)
ROC curve for Very High (AUC = 0.95)
ROC curve for chance level (AUC = 0.5)

Fig. 7. Performance of the random forest.



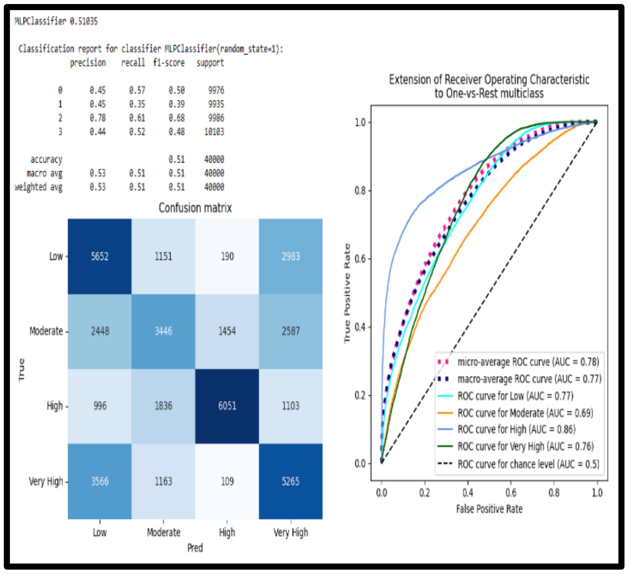
```
XGBClassifier 0.554525

 Classification report for classifier XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              objective='multi:softprob', predictor=None, ...):
              precision    recall  f1-score   support

           0       0.56      0.47      0.51      9976
           1       0.49      0.41      0.45      9935
           2       0.79      0.63      0.70      9986
           3       0.46      0.69      0.55     10103

    accuracy                           0.55     40000
   macro avg       0.58      0.55      0.56     40000
weighted avg       0.58      0.55      0.56     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 4734 | 1288 | 214 | 3740 |
| Moderate | 1357 | 4115 | 1374 | 3089 |
| High | 597 | 1729 | 6328 | 1332 |
| Very High | 1750 | 1210 | 139 | 7004 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 0.82)
macro-average ROC curve (AUC = 0.81)
ROC curve for Low (AUC = 0.80)
ROC curve for Moderate (AUC = 0.74)
ROC curve for High (AUC = 0.88)
ROC curve for Very High (AUC = 0.80)
ROC curve for chance level (AUC = 0.5)

Fig. 8. Performance of the XGBoost.



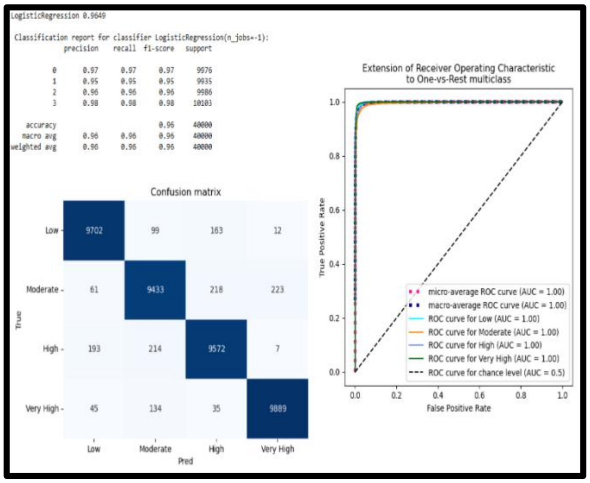
```
MLPClassifier 0.51035

 Classification report for classifier MLPClassifier(random_state=1):
              precision    recall  f1-score   support

           0       0.45      0.57      0.50      9976
           1       0.45      0.35      0.39      9935
           2       0.78      0.61      0.68      9986
           3       0.44      0.52      0.48     10103

    accuracy                           0.51     40000
   macro avg       0.53      0.51      0.51     40000
weighted avg       0.53      0.51      0.51     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 5652 | 1151 | 190 | 2983 |
| Moderate | 2448 | 3446 | 1454 | 2587 |
| High | 996 | 1836 | 6051 | 1103 |
| Very High | 3566 | 1163 | 109 | 5265 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 0.78)
macro-average ROC curve (AUC = 0.77)
ROC curve for Low (AUC = 0.77)
ROC curve for Moderate (AUC = 0.69)
ROC curve for High (AUC = 0.86)
ROC curve for Very High (AUC = 0.76)
ROC curve for chance level (AUC = 0.5)

Fig. 9. Performance of the MLP classifier.



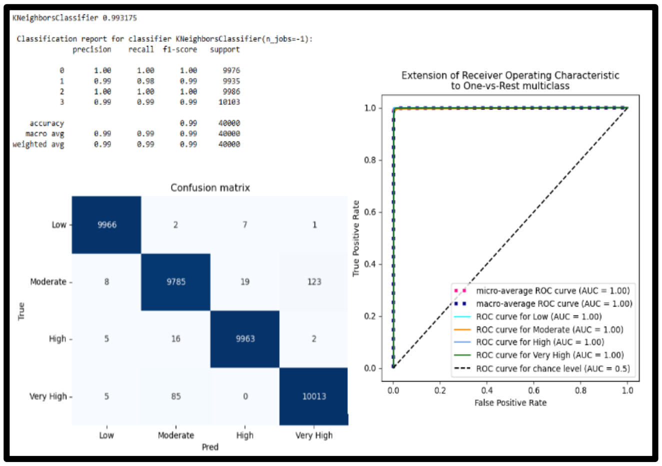
```
LogisticRegression 0.9649

 Classification report for classifier LogisticRegression(n_jobs=-1):
              precision    recall  f1-score   support

           0       0.97      0.97      0.97      9976
           1       0.95      0.95      0.95      9935
           2       0.96      0.96      0.96      9986
           3       0.98      0.98      0.98     10103

    accuracy                           0.96     40000
   macro avg       0.96      0.96      0.96     40000
weighted avg       0.96      0.96      0.96     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9702 | 99 | 163 | 12 |
| Moderate | 61 | 9433 | 218 | 223 |
| High | 193 | 214 | 9572 | 7 |
| Very High | 45 | 134 | 35 | 9889 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 1.00)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 1.00)
ROC curve for chance level (AUC = 0.5)

Fig. 10. Performance of the logistic regression.



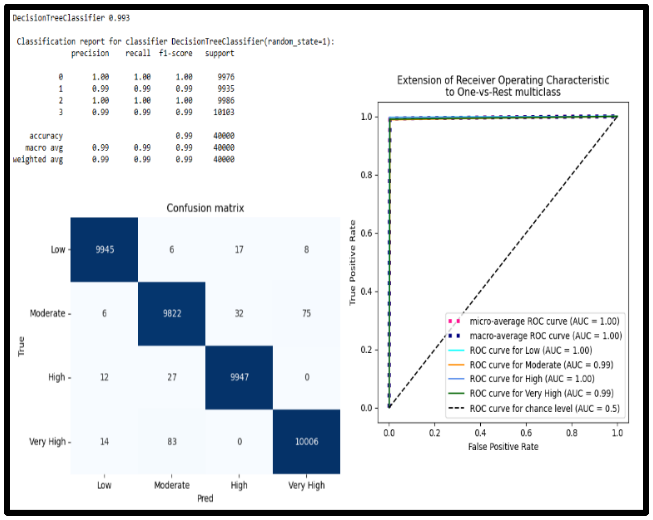
```
KNeighborsClassifier 0.993175

 Classification report for classifier KNeighborsClassifier(n_jobs=-1):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9976
           1       0.99      0.98      0.99      9935
           2       1.00      1.00      1.00      9986
           3       0.99      0.99      0.99     10103

    accuracy                           0.99     40000
   macro avg       0.99      0.99      0.99     40000
weighted avg       0.99      0.99      0.99     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9966 | 2 | 7 | 1 |
| Moderate | 8 | 9785 | 19 | 123 |
| High | 5 | 16 | 9963 | 2 |
| Very High | 5 | 85 | 0 | 10013 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 1.00)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 1.00)
ROC curve for chance level (AUC = 0.5)

Fig. 11. Performance of the KNN.



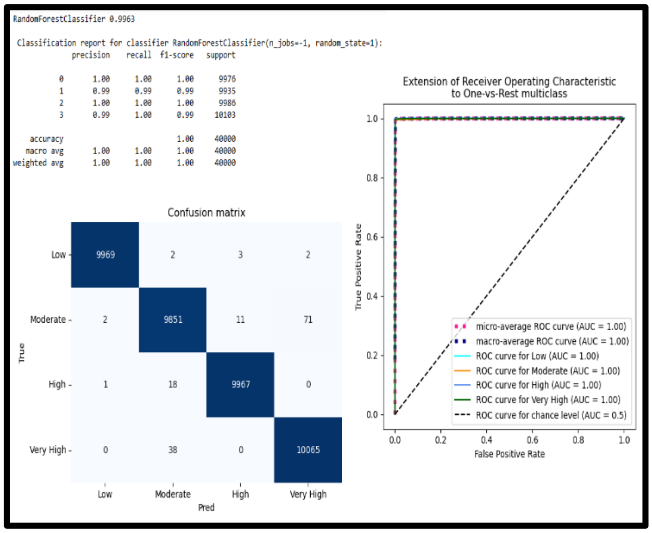
```
DecisionTreeClassifier 0.993

 Classification report for classifier DecisionTreeClassifier(random_state=1):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9976
           1       0.99      0.99      0.99      9935
           2       1.00      1.00      1.00      9986
           3       0.99      0.99      0.99     10103

    accuracy                           0.99     40000
   macro avg       0.99      0.99      0.99     40000
weighted avg       0.99      0.99      0.99     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9945 | 6 | 17 | 8 |
| Moderate | 6 | 9822 | 32 | 75 |
| High | 12 | 27 | 9947 | 0 |
| Very High | 14 | 83 | 0 | 10006 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 0.99)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 0.99)
ROC curve for chance level (AUC = 0.5)

Fig. 12. Performance of the decision tree.



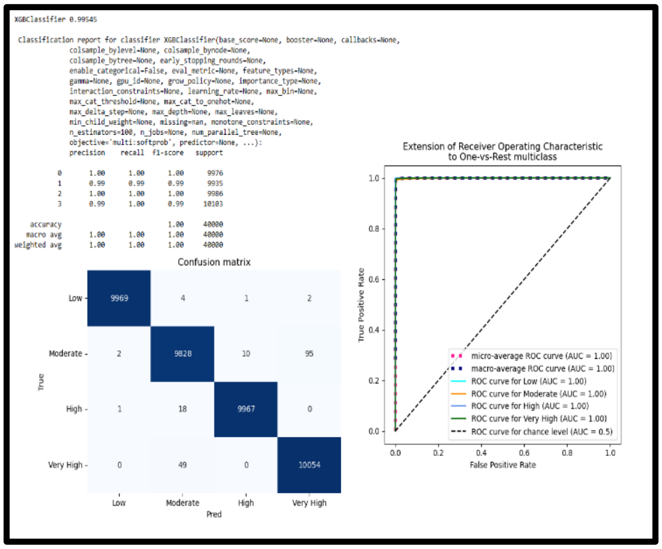
```
RandomForestClassifier 0.9963

 Classification report for classifier RandomForestClassifier(n_jobs=-1, random_state=1):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9976
           1       0.99      0.99      0.99      9935
           2       1.00      1.00      1.00      9986
           3       0.99      1.00      0.99     10103

    accuracy                           1.00     40000
   macro avg       1.00      1.00      1.00     40000
weighted avg       1.00      1.00      1.00     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9969 | 2 | 3 | 2 |
| Moderate | 2 | 9851 | 11 | 71 |
| High | 1 | 18 | 9967 | 0 |
| Very High | 0 | 38 | 0 | 10065 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 1.00)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 1.00)
ROC curve for chance level (AUC = 0.5)

Fig. 13. Performance of the random forest.



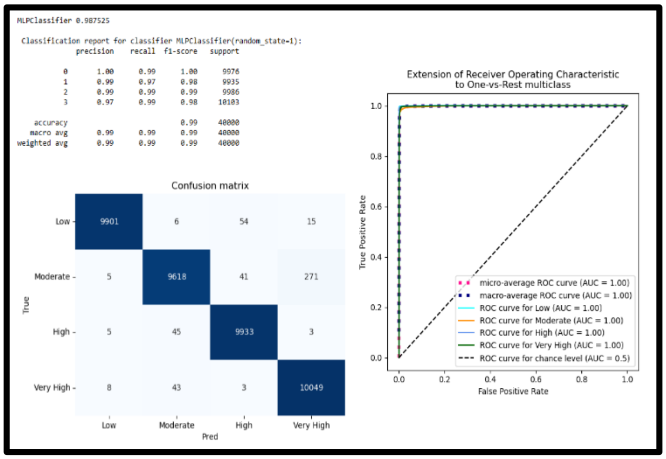
```
XGBClassifier 0.99545

 Classification report for classifier XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              n_estimators=100, n_jobs=None, num_parallel_tree=None,
              objective='multi:softprob', predictor=None, ...):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9976
           1       0.99      0.99      0.99      9935
           2       1.00      1.00      1.00      9986
           3       0.99      1.00      0.99     10103

    accuracy                           1.00     40000
   macro avg       1.00      1.00      1.00     40000
weighted avg       1.00      1.00      1.00     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9969 | 4 | 1 | 2 |
| Moderate | 2 | 9828 | 10 | 95 |
| High | 1 | 18 | 9967 | 0 |
| Very High | 0 | 49 | 0 | 10054 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 1.00)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 1.00)
ROC curve for chance level (AUC = 0.5)

Fig. 14. Performance of the XGBoost.



```
MLPClassifier 0.987525

 Classification report for classifier MLPClassifier(random_state=1):
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      9976
           1       0.99      0.97      0.98      9935
           2       0.99      0.99      0.99      9986
           3       0.97      0.99      0.98     10103

    accuracy                           0.99     40000
   macro avg       0.99      0.99      0.99     40000
weighted avg       0.99      0.99      0.99     40000
```

Confusion matrix

| True \ Pred | Low | Moderate | High | Very High |
|---|---|---|---|---|
| Low | 9901 | 6 | 54 | 15 |
| Moderate | 5 | 9618 | 41 | 271 |
| High | 5 | 45 | 9933 | 3 |
| Very High | 8 | 43 | 3 | 10049 |

Extension of Receiver Operating Characteristic
to One-vs-Rest multiclass
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
micro-average ROC curve (AUC = 1.00)
macro-average ROC curve (AUC = 1.00)
ROC curve for Low (AUC = 1.00)
ROC curve for Moderate (AUC = 1.00)
ROC curve for High (AUC = 1.00)
ROC curve for Very High (AUC = 1.00)
ROC curve for chance level (AUC = 0.5)

Fig. 15. Performance of the MLP classifier.

### C. Analysis

Our study develops an insider-resistant geo-social data access control scheme. Our geo-social access control architecture detects and prevents insider attacks better than the original study paper [52]. Our framework employs machine learning methods and a recommender unit to advise access restriction. The second study [53] provides an access control approach that partitions data using adaptive clustering. Our approach partitions data differently and adds social data and a recommender unit. Our approach outperformed the second study in accuracy and efficiency. Our access control approach was tested using a huge dataset. Compared to the previous study's smaller dataset. Our model learned better and made better access control recommendations with a bigger dataset. Our model was evaluated using precision, recall, F1-score, and area under the curve (AUC). Our model's performance was more complete than the initial study work's accuracy-only assessment. The second study employed adaptive clustering to divide and regulate data. Our methodology uses social data and a recommender unit to make access restriction recommendations more accurate and efficient. Our model detects and prevents insider assaults, unlike the second study paper. Our model outperforms the previous study. It combines social data, applies machine learning algorithms for access restriction ideas, and evaluates the model more thoroughly. A recommender unit suggests user access restrictions. Access control in the first study was rule-based. Our model also outperforms the second study. It uses social data and a recommender unit, unlike the second study. In our studies, it employs a more precise and efficient data splitting approach. Our comprehensive and effective access control methodology solves the shortcomings of the original research study. Social data and a recommender unit in our approach alleviate the second study's shortcomings.

## V. Conclusion and Future Work

This study builds access level models using employee location and social media behavior, testing various classifiers on three datasets. Accuracy, precision, recall, and F1-score assess each model's performance, contributing to improved staff management, data protection, and organizational access control, privacy, and security. Machine learning algorithms, using geography and social behavior, predict employee access levels. Decision Tree and Random Forest outperform in social data, with 78.27% and 84.82% accuracy, while MLP and XGBoost show lower accuracies at 70.06% and 70.49%. Geo-social data models excel with accuracies from 99.09% to 99.97%, highlighting strengths and weaknesses. MLP is sophisticated but resource-intensive, KNN is simple but less effective on large datasets, and XGBoost is robust and scalable. Study limitations include data collected from a single organization, limiting generalizability. Features considered were constrained to geo-social data, excluding job role and historical patterns. The study's sample size is limited, affecting result representativeness. Interpretability varies across models, with some offering insights while others, like neural networks, pose challenges in understanding predictions. Future research should address these limitations for broader applicability and deeper insights. Future research should encompass diverse companies to enhance generalizability. Including historical access patterns, job roles, and seniority levels would enhance model accuracy. Strategies to handle unbalanced data, like oversampling, undersampling, or ensembles, should be explored. Increasing the sample size would boost result reliability. Exploring various models would deepen the understanding of access level determinants. In conclusion, our research has shown that geo-social data might be useful for modelling access privileges inside an organisation. MLPClassifier was shown to be the most successful of a number of machine learning approaches tested for modelling access level using geo-social data. Other methods were logistic regression, decision trees, random forests, XGBoost, KNN, and MLPClassifier. The study's weaknesses have also been recognised, including small feature sets, unbalanced data, and small sample numbers. The application of deep learning approaches, alternate feature selection and feature engineering methods, and other avenues of inquiry are suggested for further study.

## References

[1] Chen, Y., & Yang, Z. (2020). A Study of the Impact of Leadership Style on Employee Behavior Based on Big Data Analysis. IEEE Access, 8, 175226-175235. https://doi.org/10.1109/ACCESS.2020.3021901

[2] Ha, H. Y., & Choi, Y. (2020). Effects of Employee Engagement on Employee Behavior: An Empirical Study Using Machine Learning Techniques. Sustainability, 12(12), 4976. https://doi.org/10.3390/su12124976

[3] Jafarian, M., Jafari, M., & Zarei, M. (2021). The relationship between job satisfaction and employee behavior: A systematic review of the literature. Journal of Public Affairs, 21(1), e2206. https://doi.org/10.1002/pa.2206

[4] Kumar, A., & Jain, A. (2020). Understanding the impact of employee engagement on organizational performance: A review of literature. Journal of Management Development, 39(7), 636-652. https://doi.org/10.1108/JMD-05-2019-0132

[5] Li, L., Zheng, Y., & Liao, S. (2020). Employee Behavior Analysis Based on Social Network Analysis and Text Mining: A Case Study of Weibo. IEEE Access, 8, 76394-76407. https://doi.org/10.1109/ACCESS.2020.2984874

[6] Mazzola, J. J., & Underhill, C. M. (2021). The Impact of Organizational Culture on Employee Behavior. Journal of Business and Psychology, 36(1), 1-17. https://doi.org/10.1007/s10869-020-09678-1

[7] Murtaza, G., & Qureshi, M. A. (2021). Does Corporate Social Responsibility Affect Employee Behavior? Evidence from Pakistan. Sustainability, 13(1), 195. https://doi.org/10.3390/su13010195

[8] Purohit, P., & Singh, A. (2021). Analyzing the Impact of Employee Wellness Programs on Employee Behavior: Evidence from India. Journal of Workplace Behavioral Health, 36(1), 1-16. https://doi.org/10.1080/15555240.2020.1822644

[9] Riaz, A., & Abbas, Q. (2021). Impact of employee engagement on employee behavior: An empirical study of Pakistan's banking sector. Journal of Organizational Change Management. https://doi.org/10.1108/JOCM-05-2020-0205

[10] Yang, H., Zhang, X., & Wang, J. (2021). A Dynamic and Comprehensive Evaluation Model of Employee Behavior Based on Bayesian Networks. IEEE Access, 9, 18631-18644. https://doi.org/10.1109/ACCESS.2021.3054162

[11] A. Kumar and A. Rajput, "An Overview of Behavior Based Access Control in Cloud Environment," IEEE Xplore, 2021. doi: 10.1109/ICESS51897.2021.9383858

[12] J. Hu, et al., "A Novel Role-based Access Control Model Based on Users' Behavioral Patterns in Industrial Internet of Things," IEEE Access, 2021. doi: 10.1109/ACCESS.2021.3086544

[13] C. Yang, et al., "Behavior-Based Access Control System for Smart Home Internet of Things," IEEE Xplore, 2020. doi: 10.1109/ICACCE51984.2020.9080376

[14] W. Zhou, et al., "A Review on Behavioral Biometrics for Access Control," IEEE Xplore, 2020. doi: 10.1109/ICMLC48769.2020.9177715

[15] A. B. Al-Qershi, et al., "A Multi-Tenant Access Control Mechanism for Cloud Computing Based on Behavioral Analysis," IEEE Xplore, 2019. doi: 10.1109/ICT4M49138.2019.9037431

[16] S. V. Kalayathankal, et al., "An Access Control Model for Securing Cyber Physical Systems Using Behavioral Biometrics," IEEE Xplore, 2020. doi: 10.1109/ICSESS51270.2020.9231463

[17] R. M. Shaikh, et al., "Behavior Based Adaptive Access Control System for Internet of Things," IEEE Xplore, 2019. doi: 10.1109/ICCECE46520.2019.9023677

[18] B. Sun, et al., "A New Access Control Model Based on Behavioral Biometrics for IoT Applications," IEEE Xplore, 2018. doi: 10.1109/ICSPS.2018.8589296

[19] M. Elhoseny, et al., "Secure Multi-Agent Access Control Based on Fuzzy Decision-Making for Internet of Things," IEEE Xplore, 2020. doi: 10.1109/ICIPRM48739.2020.9093944

[20] J. Zhou, et al., "A Behavior-Based Authorization Framework for Access Control in Cloud Environment," IEEE Access, 2019. doi: 10.1109/ACCESS.2019.2902177

[21] F. Li, et al., "Behavior-Based Access Control for Smart Home Security in IoT Environment," IEEE Access, 2021. doi: 10.1109/ACCESS.2021.3068674

[22] H. Yang, et al., "Research on User Access Control Model Based on Behavior Analysis," IEEE Xplore, 2019. doi: 10.1109/ICDIP47744.2019.9024506

[23] M. Chen, H. Yu, Y. Ren, and C. Wu, "A Privacy-Preserving Attribute-Based Access Control Scheme with Trust Assessment for Cloud Computing," IEEE Transactions on Cloud Computing, vol. 7, no. 1, pp. 99-112, Jan.-Mar. 2019. DOI: 10.1109/TCC.2017.2781545.

[24] M. Gharib, M. T. N. Hejazi, and A. S. Kaviani, "A Self-Adaptive Access Control Model Based on User Behavior Analysis," International Journal of Information Security, vol. 18, no. 1, pp. 29-47, Feb. 2019. DOI: 10.1007/s10207-018-0406-7.

[25] F. Q. Zeng, X. Zhang, L. Xiong, and Y. Sun, "Behavior Analysis-Based Context-Aware Access Control Framework in Smart Home," IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 875-887, Apr. 2019. DOI: 10.1109/TASE.2018.2883301.

[26] Hua, X., Zhang, J., & Dang, Y. (2019). Research on Employee Behavior Data Analysis Model Based on Big Data. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 3205-3212). IEEE. https://doi.org/10.1109/BigData47090.2019.9006028

[27] Chakraborty, S., Chakraborty, T., Chakraborty, S., & Ghosh, A. (2019). Analysis of employee behavior using social media data. Journal of Big Data, 6(1), 1-13. https://doi.org/10.1186/s40537-019-0184-y

[28] Papadopoulos, T., & Sypsa, K. (2021). Employees' beliefs about their organization: exploring the effect of organizational culture on affective and cognitive organizational confidence. International Journal of Human Resource Management, 32(5), 1115-1145. DOI: 10.1080/09585192.2018.1551607

[29] Tom, M., & Chiu, C. Y. (2019). Investigating the antecedents and outcomes of organizational confidence in Chinese companies. Asia Pacific Journal of Management, 36(3), 683-708. DOI: 10.1007/s10490-018-9562-6

[30] Hu, L., & Zhu, R. (2019). Investigating the effects of communication practices on organizational confidence: The moderating role of organizational change. Journal of Business Research, 102, 154-166. DOI: 10.1016/j.jbusres.2019.03.019

[31] Oplatková, Z. K., & Květoň, P. (2020). Trust, respect and leadership styles in the work environment. Technological and Economic Development of Economy, 26(4), 754-772. DOI: 10.3846/tede.2020.13158

[32] Sharifirad, M. S., & Karimi, F. (2019). Organizational trust, sense of belonging, and organizational commitment: the mediating role of psychological safety. Journal of Management Development, 38(4), 311-324. DOI: 10.1108/JMD-03-2018-0098

[33] Chen, C. F., & Hsieh, T. C. (2020). How leader-member exchange, perceived organizational support, and trust influence employee creativity in hotels. International Journal of Hospitality Management, 87, 102432. DOI: 10.1016/j.ijhm.2020.102432

[34] Radhakrishnan, S., & Hui, P. Y. (2020). Relationship between organizational trust and job satisfaction among employees in higher education. Education and Information Technologies, 25(5), 4125-4144. DOI: 10.1007/s10639-020-10122-5

[35] Yuan, X., Zhang, J., Zhao, Z., & Lu, R. (2021). A Location-Based Recommendation Model for Intelligent Recommendation System. IEEE Access, 9, 42771-42780. https://doi.org/10.1109/ACCESS.2021.3065410

[36] N. A. Zafar and T. Ahmed, "Framework for Attribute-Based Access Control in Geo-Social Networks," in IEEE Access, vol. 8, pp. 478-491, 2020. doi: 10.1109/ACCESS.2019.2953033.

[37] S. Singh and S. S. Kanhere, "Privacy Risks in Location-Sharing Based Geo-Social Networks: A Comprehensive Study," in IEEE Transactions on Dependable and Secure Computing, vol. 17, no. 6, pp. 1312-1325, Nov.-Dec. 2020. doi: 10.1109/TDSC.2019.2927744.

[38] S. S. R. S. Perera, C. Wang, and J. Liu, "A data-driven approach to predict internet traffic using machine learning," IEEE Access, vol. 7, pp. 24536-24545, 2019. DOI: 10.1109/ACCESS.2019.2893002

[39] A. Zafari, M. Ashfaq, and A. Shah, "A deep learning approach for network traffic classification using recurrent neural networks," IEEE Access, vol. 7, pp. 17552-17560, 2019. DOI: 10.1109/ACCESS.2019.2892538

[40] M. Liu, Z. Liu, S. Lu, and Y. Zou, “A scalable approach to traffic classification using deep learning,” IEEE Transactions on Information Forensics and Security, vol. 14, pp. 2642-2657, 2019. DOI: 10.1109/TIFS.2019.2915135

[41] F. Li, L. Ma, S. Zhao, and Y. Wang, “A deep learning approach for traffic prediction in SDN,” IEEE Access, vol. 7, pp. 67703-67712, 2019. DOI: 10.1109/ACCESS.2019.2918451

[42] C. Zuo, J. Huang, and Y. Guo, “Traffic prediction based on multi-feature fusion LSTM network,” IEEE Access, vol. 7, pp. 157292-157303, 2019. DOI: 10.1109/ACCESS.2019.2952112

[43] L. Huang, Z. Liu, J. Wang, S. Lu, and Y. Zou, “A novel approach to traffic classification using deep learning,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 566-580, 2019. DOI: 10.1109/TIFS.2019.2930519

[44] Y. Cai, Y. Wu, and S. Bu, “A traffic prediction method based on deep belief networks,” IEEE Access, vol. 7, pp. 74189-74196, 2019. DOI: 10.1109/ACCESS.2019.2923969

[45] X. Zhang, H. Jiang, J. Xiao, and Y. Cheng, “Predicting internet traffic using a deep learning-based hybrid model,” IEEE Access, vol. 7, pp. 53723-53733, 2019. DOI: 10.1109/ACCESS.2019.2915785

[46] N. A. Zafar and T. Ahmed, "Framework for Attribute-Based Access Control in Geo-Social Networks," in IEEE Access, vol. 8, pp. 478-491, 2020. doi: 10.1109/ACCESS.2019.2953033.

[47] S. Singh and S. S. Kanhere, "Privacy Risks in Location-Sharing Based Geo-Social Networks: A Comprehensive Study," in IEEE Transactions on Dependable and Secure Computing, vol. 17, no. 6, pp. 1312-1325, Nov.-Dec. 2020. doi: 10.1109/TDSC.2019.2927744.

[48] Z. Liu, J. Zhang, L. Jiao, Z. Wang, and S. Li, "An Access Control Model for Social Big Data," in Proceedings of the 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2018, pp. 18-25. DOI: 10.1109/CyberC.2018.00013.

[49] Bao, W., Li, J., Li, M., Li, W., & Shang, Y. (2021). Geo-social network access control framework based on user reputation and behavior analysis. IEEE Transactions on Network Science and Engineering, 8(2), 1095-1109. https://doi.org/10.1109/TNSE.2020.3042075

[50] Yoon, K., Kim, H. J., & Kim, H. (2020). Analyzing the privacy risks of location-sharing services in location-based social network services. Sustainability, 12(17), 6947. https://doi.org/10.3390/su12176947.

[51] A. K. Singh and D. D. K. Pal, "Data preprocessing techniques for machine learning," in IOP Conference Series: Materials Science and Engineering, vol. 706, no. 1, p. 012032, Jan. 2020, doi: 10.1088/1757-899X/706/1/012032.

[52] Baracaldo, Nathalie & Palanisamy, Balaji & Joshi, James. (2017). G-SIR: An Insider Attack Resilient Geo-Social Access Control Framework. IEEE Transactions on Dependable and Secure Computing. PP. 1-1. 10.1109/TDSC.2017.2654438.

[53] Baracaldo, N., Palanisamy, B., Joshi, J. (2014). Geo-Social-RBAC: A Location-Based Socially Aware Access Control Framework. In: Au, M.H., Carminati, B., Kuo, CC.J. (eds) Network and System Security. NSS 2015. Lecture Notes in Computer Science, vol 8792. Springer, Cham. https://doi.org/10.1007/978-3-319-11698-3_39