



## PREDICTION OF CLIENT SATISFACTION LEVELS WITH GHANA'S LAND ADMINISTRATION SYSTEM: USING A MACHINE LEARNING APPROACH

<sup>1</sup>Tabase, F. <sup>2</sup>Kumi-Boateng, B. & <sup>1</sup>Yakubu, I.

<sup>1</sup>University of Mines and Technology (UMaT), School of Railways and Infrastructure Development (SRID),  
Geomatic and Civil Engineering Department, P.O. Box 237, Tarkwa, Ghana.

<sup>2</sup>University of Mines and Technology, Geomatic Engineering Department, Faculty of Geoscience and  
Environmental Studies

Correspondence Author: [ftabase@umat.edu.gh](mailto:ftabase@umat.edu.gh)

### Abstract

Ghana's first digital Land Administration System (LAS), known as the Client Service Access Unit (CSAU) managed by the Lands Commission (LC) is responsible for the provision of all land documentation procedures and is the one-stop shop for land transactions. However, since its creation as a result of Ghana's Land Administration Projects (LAP 1 and LAP 2), there is no system available to determine the satisfaction levels of clients to maintain client loyalty to the system. This study aims to develop predictive models to forecast client satisfaction (CS) with the CSAU using machine learning (ML) algorithms including Random Forest (RF), Decision Tree (DT), Naive Bayes (NB) and Support Vector Machine (SVM) using Jupyter Notebook Software and to determine the most important factors influencing client satisfaction with the services rendered by the CSAU, based on 199 responses. Among the ML algorithms, Random Forest and Decision Tree had the highest predictive accuracy of 85%, and among the eleven (11) latent factors used in the study, Value for Money had the strongest influence on CS with the services provided by the CSAU. These findings provide the basics of incorporating machine learning algorithms into real-time client satisfaction analysis dashboards to enable the service providers of the CSAU to identify areas for improvement and the overall satisfaction of clients with the services they provide. This study further contributes to the general world of research on leveraging advanced ML algorithms and Artificial Intelligence (AI) to solve problems in the field of service delivery assessment and client satisfaction.

**Keywords:** Land Administration Systems (LAS), Client Service Access Unit (CSAU), Client Satisfaction (CS), Classification, Machine Learning (ML) Algorithm

### Introduction

Effective land management is achieved through the use of a well-planned and regulated LAS. This is a very important aspect of a nation's development as it ensures land tenure security, appropriate land use planning and reduces land-related disputes and conflicts (Indrajit et al., 2021). However, many developing countries, including Ghana, face challenges with their land resource management due to the lack of a well-planned and regulated LAS (Ameyaw & de Vries, 2021; Ahey, 2023). While Ghana over the past decade has had

many policies and implementation strategies to boost its land management sector and to bring certainty to rights and interests to users, which includes projects such as the Land Administration Project phases 1 and 2 (LAP 1 and LAP 2, respectively), which was made possible by the introduction of the National Land Policy (NLP) of 1999 (Alhola & Gwaindepi, 2024). The LAP 1 focused on the decentralisation of the LAS services to regional levels of the LC and the introduction of the Customary Land

Secretariat (CLS) and the LAP 2 focused on the establishment of Ghana's first Land Information System (LIS) called the Ghana Enterprise Land Information System (GELIS) which metamorphosed into the Enterprise Land Information System (ELIS) and now piloted only in Greater Accra Region (Owusu Ansah et al., 2024). LAP 2 also contributed to the establishment of Ghana's sole digital LAS called the CSAU, usually termed as the one-stop-shop for all land services in Ghana, including deeds/title registration, official/certificate search, cadastral plan approval, payment of ground rent, general valuation and many more.

While these initiatives have led to the development of Ghana's LAS, the CSAU is bedevilled with challenges, including inefficient documentation processes, corruption, lack of transparency and delays in service delivery which continue to undermine public confidence in the benefits of the system (Schoneveld & German, 2013), hence the need to develop a ML technique to model and predict CS of the CSAU in order to forecast the levels of satisfaction derived by clients and to determine which of the various latent factors influence satisfaction the most in the CSAU of the LC.

Shimizu and Kaneko (2021), proposed a hybrid model made of Decision Trees (DT) and Random Forest (RF), which provides a high prediction accuracy and interpretability, considering the strength of DT for accurate global interpretation of data and the ability of RF to improve prediction accuracy and local interpretations, but did not take into consideration other models, including NB and SVM to further compare to broaden the search for a more perfect prediction model. Tan et al. (2024), compared the RF and SVM to obtain the best prediction model regarding land classification, and it was concluded that the SVM has a superior prediction accuracy than that of the RF but

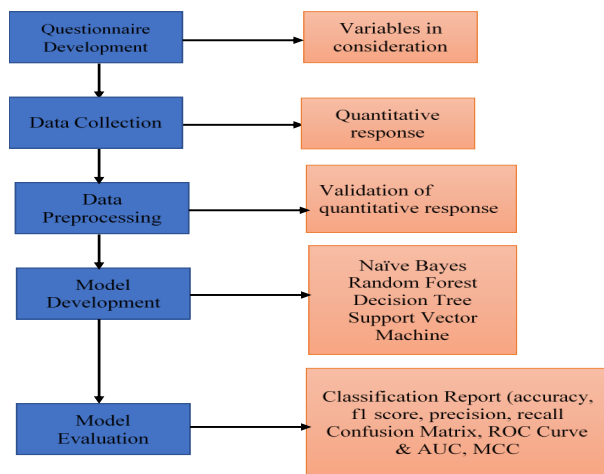
did not take into consideration other models, such as the DT and NB to broaden the scope for the search for the best prediction model for land classes. Mutale et al.(2024), evaluated the performance of RF, SVM and Artificial Neural Network (ANN) in the prediction of land use- land cover dynamics, and it was concluded that the RF model outperformed the SVM and ANN models but did not take into consideration the NB model and these models were not applied in prediction of CS. Jijo and Abdulazeez (2021), applied the DT algorithm for classification, but failed to use other models such as the NB, SVM and the RF for the classification. Halifa and Novita (2023), applied the NB algorithm for the prediction of CS with the Post Office Services of Rumbai and concluded that the NB algorithm performed excellently in the prediction of CS, but this study failed to use other algorithms, such as the SVM, DT and RF, to compare their performance with that of the NB to obtain the best performing model.

This study therefore aims to develop predictive models to forecast CS with the CSAU of the LC of Ghana using the Random Forest, Decision Tree, Naive Bayes and Support Vector Machine and to determine which latent factor significantly influences client satisfaction with the CSAU and also determine the best performing machine learning algorithm to be incorporated in the CSAU CS application to give real-time feedback on satisfaction levels of clients on daily basis.

### **Methodology**

Figure 1 shows the flowchart of the methodology for this study, which followed a five-stage approach, starting from the development of the questionnaire to obtaining respondent data, to evaluating the machine learning models used for this study.

**Figure 1: Diagram of Research Methodology**



*Source: Author’s Construct (2025).*

**Questionnaire Development and Data Collection**

The data collection for this study took into consideration the development of the questionnaire, the dissemination of the

questionnaire and the sorting of the data obtained to meet the format required for processing in Jupyter NoteBook Software for model development and to determine the best performing model for CS. The factors considered in the questionnaire were selected through a literature review, Land Administration Expert advice, and inputs from academia in the area of Land Administration. The targeted respondents for the questionnaire were the clients of the CSAU of the Lands Commission of Ghana for all sixteen (16) regions using the KoboToolbox software.

**Causal Factors Considered**

To assess client satisfaction, the factors influencing CS must be assessed; therefore, twelve factors were considered in this study, including overall customer satisfaction. Table 1 expresses these factors and their respective symbols, which will represent them throughout the study.

**Table 1: Causal factors considered**

| l | Causal Factor                 | Symbol | Causal Factor                     |
|---|-------------------------------|--------|-----------------------------------|
|   | Overall Customer Satisfaction | VM     | Value for Money                   |
|   | Security                      | AnR    | Awareness and Reception           |
|   | Safety                        | ACS    | Accessibility and Internet Issues |
|   | Service Predictability        | T      | Transparency                      |
|   | Cost of service               | PU     | Progress Updates                  |
|   | Service Delivery              | PI     | Power Issues                      |

*Source: Author’s Construct (2025).*

**Data Pre-processing**

The data pre-processing phases included handling missing data in the respondent feedback obtained. This was handled through the median imputation method rather than deletion, except in cases where a significant portion of responses were missing.

The next step of the data pre-processing was the Likert scale reclassification, while the questionnaire followed the approved five-point Likert scale rule to enable better representation. For better machine learning assessment, respondent feedback was reclassified to a simple three-point Likert scale as shown in Table 2.

**Table 2: Likert Scale Reclassification for Customer Satisfaction**

| Original Likert Scale | Response Label    | Reclassified Class |
|-----------------------|-------------------|--------------------|
| 1                     | Very Satisfied    | 1 (Satisfied)      |
| 2                     | Satisfied         | 1 (Satisfied)      |
| 3                     | Neutral           | 2 (Neutral)        |
| 4                     | Dissatisfied      | 3 (Unsatisfied)    |
| 5                     | Very Dissatisfied | 3 (Unsatisfied)    |

*Source: Author’s Construct (2025).*

The last stage of the data pre-processing involved randomly splitting the reclassified data into training and testing datasets; an 80% training and 20% testing split was done in Jupyter NoteBook Software.

**Model Development**

The model development phase involved training each of the ML algorithms on the training data and evaluating the final performance on the testing data in Jupyter NoteBook Software. Therefore, the study used a supervised machine learning approach.

**Model Validation**

The last stage of the study is the model validation; the performance of each trained machine learning model was evaluated after its prediction with the testing data set. Chicco and Jurman (2023) proposed that Matthews Correlation Coefficient (MCC) should replace the Area Under Curve (AUC) and recall as a standard metric for assessing binary classification, since recall and AUC do not show precision and negative predictive values from the Confusion Matrix (CM), while MCC does not use all the cells of the CM. Since this study is not considering a binary class of latent factors but a multi-class of latent factors, all the performance metrics are applied. The evaluation matrices used for this study included accuracy, precision, recall, F1-score, MCC and AUC in Jupyter NoteBook Software.

**Results**

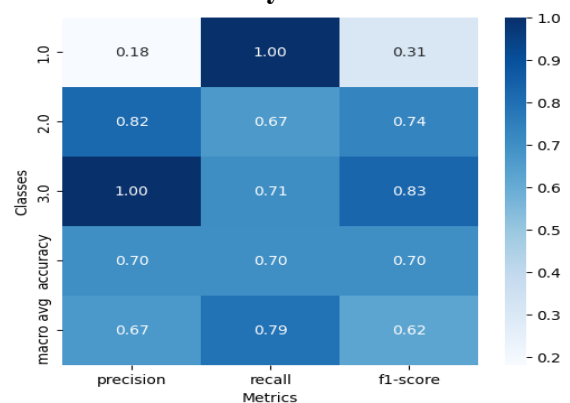
**Classification Reports**

The classification reports of NB, DT, RF and SVM are expressed in the subsequent subsections under results.

*Naïve Bayes*

The NB model reached an accuracy of 70%, the MCC was 0.56 and 0.84 micro-average AUC. The macro averages precision, recall, F1-score were 0.67, 0.79, and 0.62, respectively, whereas the precision, recall and the F1-score of the weighted averages that consider class support were 0.87, 0.70, and 0.75, respectively. The findings show that although the model performs well in class 3 precision, it has difficulties with class 1 precision and class 2 recall. The results are presented in Figure 2.

**Figure 2: Classification Report of the Predicted Naïve Bayes Model**

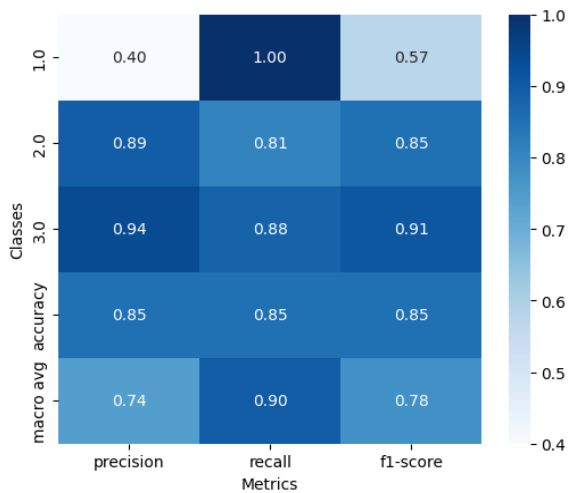


*Source: Author’s Construct (2025)*  
*Decision Tree*

The DT achieved an accuracy of 85%, the MCC was 0.75 and 0.89 micro-average AUC. The macro averages precision, recall, F1-score were 0.74, 0.90, and 0.78, respectively, whereas the precision, recall and the F1-score of the weighted averages that consider class support were 0.89, 0.86, and 0.86, respectively.

The performance was strong across all classes, with significant improvement in Class 1 F1-score. The results are illustrated in Figure 3.

**Figure 3: Classification Report of the Predicted Decision Tree Model**

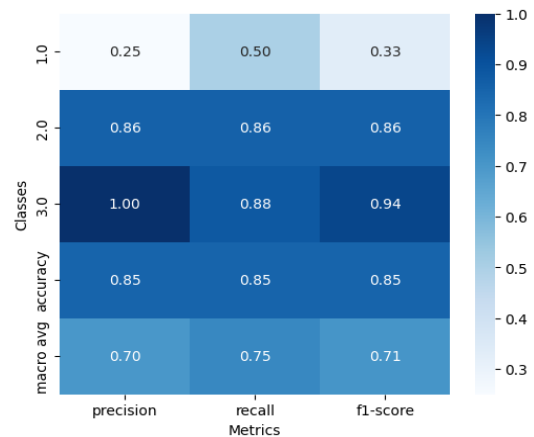


*Source: Author's Construct (2025).*

**Random Forest**

The RF model achieved an accuracy of 85%, an MCC of 0.75 and a 0.97 micro-average AUC. The macro averages precision, recall, F1-score were 0.70, 0.75, and 0.71, respectively, whereas the precision, recall and the F1-score of the weighted averages were 0.89, 0.85, and 0.87, respectively. The performance was strong for Classes 2 and 3. The results are shown in Figure 4.

**Figure 4: Classification Report of the Predicted Random Forest Model**

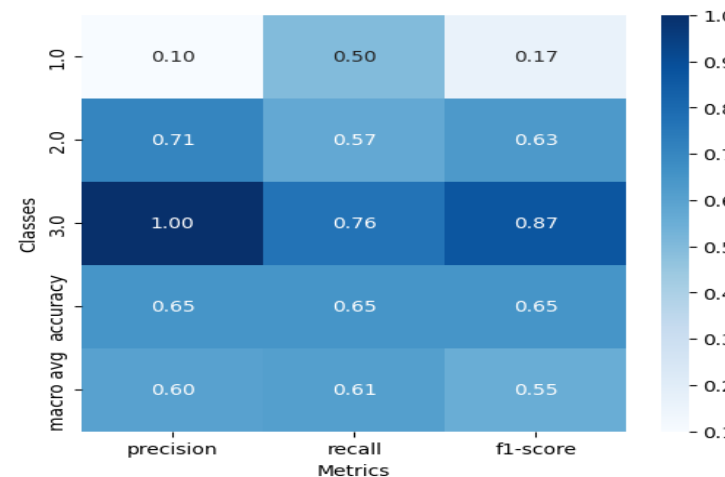


*Source: Author's Construct (2025).*

**Support Vector Machine**

The SVM achieved the lowest accuracy of 65%. The macro averages of the precision were 0.60, the recall of 0.61, and the F1-score of 0.55, while the precision, recall and the F1-score of the weighted averages showed 0.80, 0.65, and 0.71, respectively. While Class 3 predictions were excellent, the poor results for Class 1 and the lower recall for Class 2 reduced the overall performance. The results are shown in Figure 5.

**Figure 5: Classification Report of the Predicted**



*Source: Author's Construct (2025).*

**Model Evaluation**

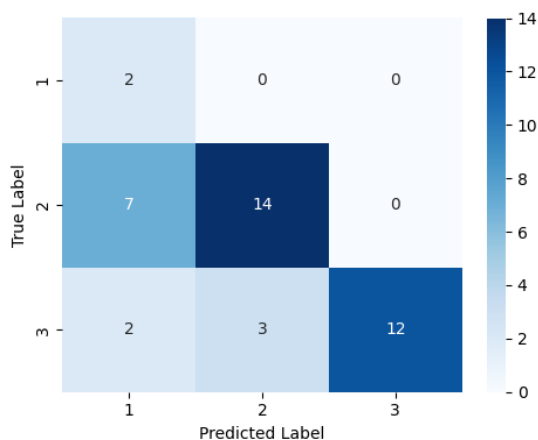
The confusion matrix was computed for the number of instances that were

correctly classified and misclassified for each class category in Jupyter Notebook Software. This helps visualise the model's ability to differentiate between different classes and shows any misclassifications by providing insights where the model might exhibit bias. The Confusion Matrices of NB, DT, RF and SVM are expressed in the preceding subsections.

### Naïve Bayes

The NB performance for class 1 correctly predicted actual class 1 instances as class 1, with several instances from other classes being misclassified as class 1, explaining the low precision of 0.18. Most Class 2 instances were correctly identified, 14 out of 21, with a few instances misclassified as Class 2 and 7 instances of Class 2 were predicted as other classes. For class 3, the model correctly classified 12 of 17 class 3 instances with no misclassification from other classes, and only 5 class 3 cases were predicted as other classes. The finding shows that the model performed well for class 3 precision and class 1 recall, and misclassification occurred mostly between classes 2 and 3. The results are shown in Figure 6.

**Figure 6: Confusion Matrix of the Predicted Naïve Bayes Model**

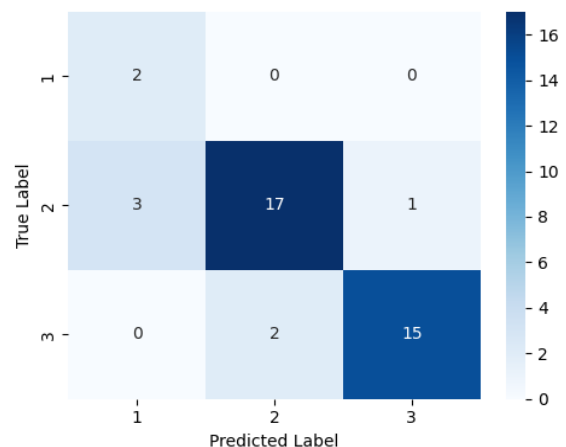


*Source: Author's Construct (2025).*

### Decision Tree

The DT correctly identified both actual class 1 instances, were correctly identified with 1 instance from another class misclassified as class 1 and a full recall for class 1 false negatives. For class 2, 17 out of 21 cases were correctly classified, with a small number of cases from other classes incorrectly labelled as class 2 and 4 Class 2 instances were misclassified as other classes. 15 of 17 cases were correctly classified for class 3, with 2 Class 3 instances predicted as other classes. The results show a strong recall for class1 and a high accuracy of class 3, and most misclassification happening between class 2 and 3. The results are shown in Figure 7.

**Figure 7: Confusion Matrix of the Predicted Decision Tree Model**



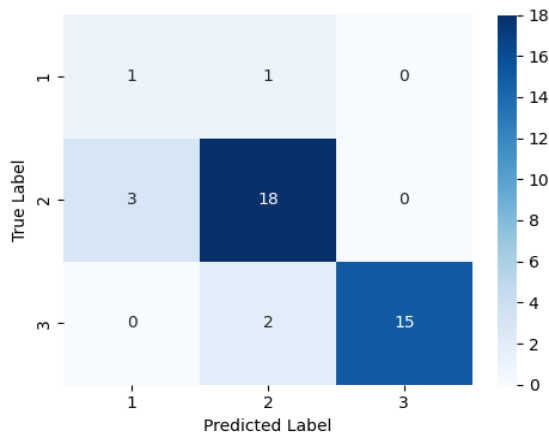
*Source: Author's Construct (2025).*

### Random Forest

The RF performed 1 out of 2 for true positives for class 1 instance with a few instances from other classes misclassified as class 1, and only 1 class 1 case predicted as another class. For class 2, 18 of 21 were correctly identified, with some misclassified false positives and 3 Class 2 instances predicted incorrectly. For class 3, 15 of 17 cases were correctly classified, and 2 class 3 cases were predicted as other classes. The Random Forest indicates excellent classification

of class 2 and class 3, with minor recall issues for class 1, shown in Figure 8.

**Figure 8: Confusion Matrix of the Predicted Random Forest Model**

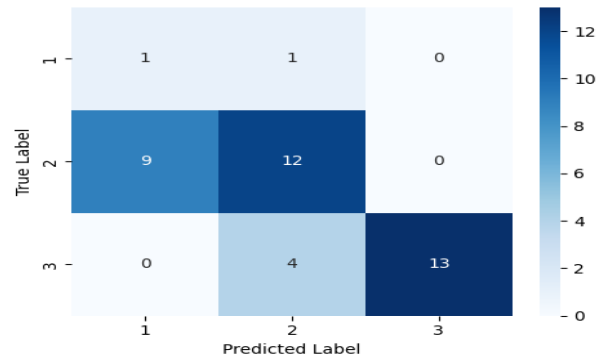


*Source: Author's Construct (2025).*

**Support Vector Machine**

SVM class 1 recorded 1 of 2 cases correctly classified, with several instances from other classes predicted as class 1, the low precision score of 0.10 and 1 class 1 case was predicted as another class. 12 of 21 class 2 cases were correctly classified, with 9 class 2 instances predicted as other classes, lowering the recall to 0.57. The class 3 instances correctly predicted 13 out of 17, with no false positive and 4 Class 3 instances were misclassified. The SVM confusion matrix shows that while class 3 predictions were highly accurate, Class 1 precision was poor and low recall from class 2 is illustrated in Figure 9.

**Figure 9: Confusion Matrix of the Predicted Support Vector Machine Model**



*Source: Author's Construct (2025).*

**Model Comparison**

DT and RF achieved the highest accuracy of 85%, followed by NB and SVM of 65% and 70%, respectively.

The DT achieved the highest MCC of 0.75, indicating the most robust overall classification performance across all four models. The RF also achieved an AUC of 0.97, a MCC of 0.74 and an accuracy of 0.85. This shows a good model performance, slightly above that of the DT. The SVM had the worst model performance by accuracy, with an accuracy of 0.65 and a MCC of 0.47; the NB performed slightly higher than the SVM, with an accuracy of 0.70 and a MCC of 0.56.

Overall, the best model by performance was the DT and RF, with both having an accuracy of 0.85.

**Table 3: Comparison of the Four Models**

| Model         | Accuracy | MCC      | auc_micro |
|---------------|----------|----------|-----------|
| Naive Bayes   | 0.70     | 0.564221 | 0.844687  |
| Decision Tree | 0.85     | 0.745466 | 0.887500  |
| Random Forest | 0.85     | 0.735739 | 0.972656  |
| SVM           | 0.65     | 0.465296 | 0.935938  |

*Source: Author's Construct (2025).*

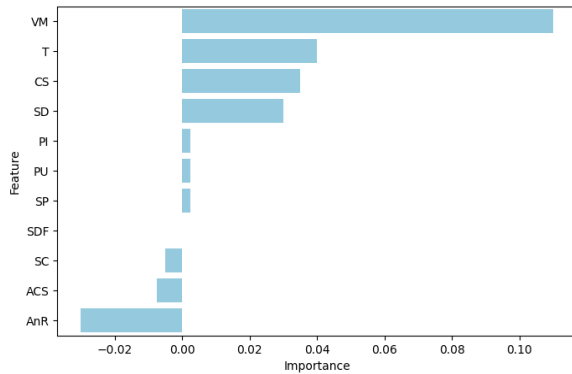
**Feature and Permutation Importance**

Figure 10 shows the factors that had the most importance to CS by the NB model.

Value for Money was the most important feature for predicting client satisfaction,

which indicates that the CSAU should focus on the cost of services provided for its clients and whether the services are worth the price. This is followed by transparency, security and service delivery.

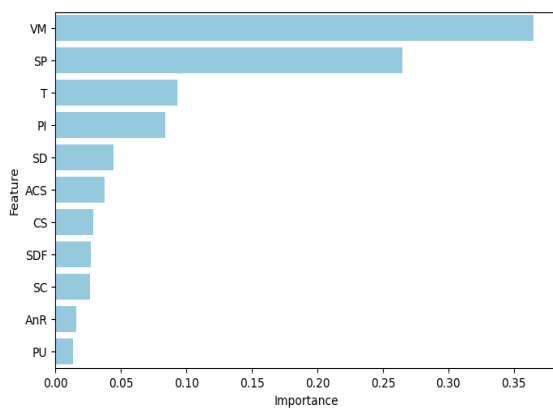
**Figure 10: Permutation Importance for Naive Bayes**



*Source: Author's Construct (2025).*

Figure 11 shows that Value for money again is the most important factor for predicting client satisfaction from the DT model, followed by service predictability and transparency. This again points to the fact that the CSAU must pay critical attention to its cost of service.

**Figure 11: Feature Importance for Decision Tree**

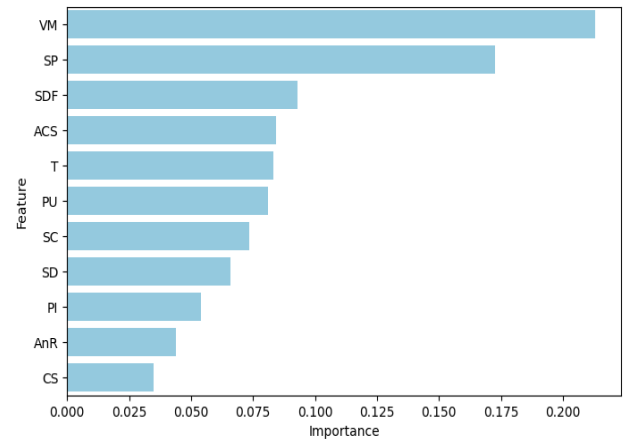


*Source: Author's Construct (2025).*

Similarly, in Figures 12 and 13, the feature importance for RF and SVM, respectively, value for money once more had the highest score and is the most important predictor of CS. This shows the importance of the

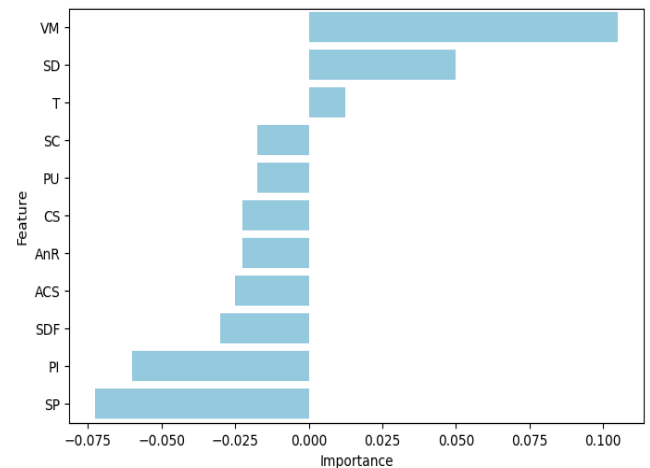
cost of service as influencing client behaviour.

**Figure 12: Feature Importance for Random Forest**



*Source: Author's Construct (2025).*

**Figure 13: Permutation Importance for SVM**



*Source: Author's Construct (2025).*

### Discussion

The findings of this study revealed that the RF and DT were the best ML models to be used in the prediction of CS in a real-time client satisfaction analysis dashboard for the CSAU. They both achieved an accuracy of 0.85, which means the prediction ability of the models were 85% correct using the testing dataset on the trained model. Both models achieved such predictive power that the client satisfaction software could integrate both models to complement each other. The NB model was the third-best-performing

model with an accuracy of 0.70, which means its predictive power was 70%; a 30% error in prediction makes it less advisable for integration into a client satisfaction software; it could be added to the best-performing models to create a more complex system. The SVM was the less performing model with an accuracy of 0.65, while the SVM is a robust machine learning model, its predictive power for CSAU client satisfaction is comparatively low.

Also, with the feature importance of the factors affecting CS, Value for Money was the most important factor in all four models. This indicates that the cost of services at the CSAU is a very strong predictor of how clients perceive satisfaction; the CSAU must pay critical attention to their service costs. Transparency, Service Predictability and Service delivery were also highly influential factors for predicting client satisfaction.

### Conclusion

It can be concluded that the Decision Tree and Random Forest machine learning models are the best algorithms to be used in a real-time client satisfaction dashboard, as they both obtained a predictive accuracy of 85%, and will complement each other if incorporated in the dashboard to predict client satisfaction with the CSAU services in real time.

Finally, it can be concluded that the most important factor in predicting client satisfaction is Value for Money, which indicates that the cost associated with the quality of services provided by the CSAU has the strongest effect on a client's perception of satisfaction with the service being rendered. Other factors which had high predictive effects are Transparency, Service Delivery and Security, which must not be avoided when dealing with the assessment of CS of land administration service delivery.

### Recommendation

It is therefore recommended that the Lands Commission of Ghana should develop a computer application based on the results obtained from this study to receive inputs of users and clients of the CSAU on daily basis at all CSAU offices in Ghana to collate a database (which is currently non-existent) for the measurement and prediction of satisfaction levels of CSAU in real-time and for the future using a hybrid of Decision Trees and Support Vector Machines algorithms. This will serve as the simplest and most transparent approach to measuring the success of the CSAU, currently in a pilot stage and will offer solutions to the full and effective implementation in all the regions.

### References

- Ahey, M. D. (2023). *Challenges of land administration in Ghana: A review of the Ghana national land policy* (FIG Working Week 2023: Protecting our world, conquering new frontiers) Orlando, Florida, USA, 28 May–1 June 2023. [https://www.fig.net/resources/proceedings/fig\\_proceedings/fig\\_2023/papers/cinema11/CINEMA11\\_ahey\\_11919\\_abs.pdf](https://www.fig.net/resources/proceedings/fig_proceedings/fig_2023/papers/cinema11/CINEMA11_ahey_11919_abs.pdf)
- Alhola, S. & Gwaindepi, A. (2024). Land tenure formalization and perceived tenure security: Two decades of the land administration project in Ghana, *Land Use Policy*, 143, 107195. <https://doi.org/10.1016/j.landusopol.2024.107195>
- Ameyaw, P. D. & de Vries, W. T. (2021). Toward smart land management: Land acquisition and the associated challenges in Ghana. a look into a blockchain digital land registry for prospects. *Land*, 10(3), 239. <https://doi.org/10.3390/land10030239>

- Chicco, D. & Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16(4). <https://doi.org/10.1186/s13040-023-00322-4>
- Halifa, A. & Novita, R. (2023). Application of naïve bayes classifier algorithm in determining the level of customer satisfaction with Rumbai post office services. *Jurnal Teknik Informatika (JUTIF)*, 4(6), 1295–1304. <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1054>
- Indrajit, A., Van Loenen, B., Suprajaka, N., Jaya, V. E., Ploeger, H., Lemmen, C., & Van Oosterom, P. (2021). Implementation of the spatial plan information package for improving the ease of doing business in Indonesian cities. *Land Use Policy*, 105, 105338: <https://doi.org/10.1016/j.landusepol.2021.105338>
- Jijo, B. T. & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(1), 20-28. <https://doi.org/10.38094/jastt20165>
- Mutale, B., Withanage, N. C., Mishra, P. K., Shen, J., Abdelrahman, K., & Fnais, M. S. (2024). A performance evaluation of random forest, artificial neural network, and support vector machine learning algorithms to predict spatio-temporal land use-land cover dynamics: A case from Lusaka and Colombo. *Frontiers in Environmental Science*, 12, 1431645. <https://doi.org/10.3389/fenvs.2024.1431645>
- Owusu Ansah, R., Abubakari, Z., Quaye, B., Todorovski, D., Malumbo, C., & Lemmen, C. (2024). Navigating the terrain of digital transition: Ghana's journey of developing a digital land information system. *Land*, 13, 528. <https://doi.org/10.3390/land13040528>
- Schoneveld, G. C., & German, L. (2013). Translating legal rights into tenure security: Lessons from the new commercial pressures on land in Ghana. *The Journal of Development Studies*, 50(2), 187–203. <https://doi.org/10.1080/00220388.2013.858129>
- Shimizu, N. & Kaneko, H. (2021). Constructing regression models with high prediction accuracy and interpretability based on decision tree and random forests. *Journal of Computer Chemistry, Japan*, 20(2): 71–87. <https://doi.org/10.2477/jccj.2020-0021>
- Tan, Y.-C., Duarte, L., & Teodoro, A. C. (2024). Comparative study of random forest and support vector machine for land cover classification and post-wildfire change detection. *Land*, 13(11), 1878. <https://doi.org/10.3390/land13111878>