International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 01 – 04, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1001

App Sheet for Course Attendance Monitoring in Higher Learning Institutions: A Citizen Developer Approach

Kelvin Rweshobora
Eastern Africa Statistical Training Centre
Dar es salaam, Tanzania

Ali Khelef
Eastern Africa Statistical Training Centre
Dar es salaam, Tanzania

Ramadhani Mrisho
Eastern Africa Statistical Training Centre
Dar es salaam, Tanzania

Abstract: Low-code and no-code development platforms have been around for a long time, but they have only recently become popular. Platforms like Appian and Mendix make users with no coding experience to build complex applications. Citizen developers have been on the rise since the inception of low-code/ no-code app development. However, this phenomenon has not yet benefited learning institutions in Tanzania. Many of the tasks performed by instructors are repetitive and time-consuming, such tasks include recording and tracking student attendance.
A traditional approach to record student attendance is performed by asking every student to sign on an attendance list that passes through all students during the beginning of lectures. This approach is inefficient in term of time and reliability especially in a large class, where a student can sign on the attendance list for other students who are not present in the class. Several automated attendance systems have been proposed based on biometric recognition, barcode, and other technologies such as mobile device. However, all of the previous developed systems require programming skills which sometimes become difficult for Citizen developer to come up with such system.
AppSheet as a no-code development platform provides features that empowers non-developers to customize their applications to suit their specific organizational needs. It enables users to create intelligent actions and interfaces autonomously. Thus, this paper explore the potential of App Sheet by proposing the Attendance Monitoring Application (CAMA) to automate student attendance processes for learning institutions in Tanzania. The results have shown that the CAMA can be used as a solution to assist course managers in registraton and monitoring of students perticipation in the course.

Keywords: App Sheet, low-code, no-code, business process automation, Citizen Developer

1. INTRODUCTION

Low-code and no-code development platforms have been around for a long time, but they have only recently become popular. McLean (2021) defines low/no-code development as a visual software development method that allows developers to build mobile or web applications by dragging and dropping components. In the early 2000s, low-code platforms became more popular, with platforms like Appian and Mendix providing users with a graphical user interface for building applications. These platforms made it possible for users with no coding experience to build complex applications. The COVID-19 pandemic has contributed the rise and awareness of no/low code development platforms. This in turn has paved way for the rise of citizen developers who have proved to be able to create useful and department specific software solutions. Citizen developers have been on the rise since the inception of low-code, no-code app development. However, this phenomenon has not yet benefited learning institutions in Tanzania.

Many of the tasks performed by instuctors are repetitive and time-consuming, such tasks include recording and tracking student attendance. Student attendance is a crucial factor for the performance of the students. In most of the academic institutions, student attendance is taken as one of the requirements for the student to take the exam (Sunaryono et al., 2021). A traditional approach to record student attendance is performed by asking every student to sign on an attendance list that passes through all students during the beginning of lectures. This approach is inefficient in term of time and reliability especially in a large class, where a student can sign on the attendance list for other students who are not present in the class. Several automated attendance systems have been proposed based on biometric recognition, barcode, and other technologies such as mobile device (Mahboob et al., 2016; Sunaryono et al., 2021). However, all of the previous developed systems require programming skills which sometimes become difficult for non-programmer (Citizen developer) to come up with such system.

AppSheet is an application that provides a no-code development platform for application software using data sources like Google Drive and other cloud-based spreadsheet and database platforms(“AppSheet,” 2023). It provides opportunity for citizen developer such as course instructor to develop an application that can automate the student attendance. AppSheet simplifies the development process by allowing simultaneous creation of desktop and mobile applications with minimal coding required. This flexibility empowers non-developers to customize their applications to suit their specific organizational needs. Additionally, AppSheet incorporates artificial intelligence and utilizes Google's hosting language, enabling users to create intelligent actions and interfaces autonomously. Thus, this paper explore the potential of App Sheet by proposing the Attendance Monitorng Application (CAMA) to automate student attendance processes for learning institutions in Tanzania.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 01 – 04, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1001

2. RELATED WORK

Several organizations have seen the potential for using low-code platforms for web system and app development. Wijesekara et al., (2020) developed an electronic record keeping system at a pediatric clinic in Colombo South Teaching Hospital, Sri Lanka with the aim of obtaining a user friendly electronic DBMS for record keeping, audit and data analysis.The group used googlesheet to create data table as the backend which was then linked to AppSheet for the front end component of the app.

(Martinez & Cisterna, 2023) Petrović et al.,(2020) showcased the application of AppSheet as they demonstrated a rapid development of data driven application using Appsheet and Apps Script. They were able to come up with two artifacts from two use cases, smart home energy management application and COVID-19 patient risk assessment. They have achieved significant development time reduction as compared to traditional app development approach.

The construction industry has also seen the application of low-code in the digitization of its daily processes with the aim of improving the industry.The authors in (Martinez & Cisterna, 2023) discuss the development and deployment of a low-code/AI solution that automates the data processing of paper delivery notes on-site .They observed a significant change in the processing time in comparison to their previous process.The low-code powered process showed a 78% reduction in process time.

The Ponorogo Regency library and Archives Office have also utilized the services of AppSheet as a means to optimize the performance of Archivist Human Resources. From the results AppSheet improved the effectiveness of the performance of archivist human resources in the Office of the Library and Archives of Ponorogo Regency (Sitaviana & Indrahti, 2022).

3. CASE STUDY

Teaching process covers many aspects like attendance tracking, record keeping, continuous assesment and report production. These activities are both tidias and time consuming. with the aid of low-code, this paper aims to empower teachers in every levelof the education system with developers power to automate these tasks. They will assume the role of citizen developers.
This paper showcases a low-code applications Course Attendance Monitoring Application (CAMA) developed using Googles AppSheet and GoogleSheet. This application focuses on course Attendance Monitoring enabling proper record keeping and monitoring of students perticipation.

4. METHODOLOGY

In this context, we have chosen rapid application development (RAD) as the primary approach due to its adaptability in quickly creating an initial working prototype of the application and subsequently refining it continuously.

4.1 APPLICATION ARCHITECTURE

The designing of CAMA involves three google sheet tables for data storage and Appsheet for the UI design figure 1.

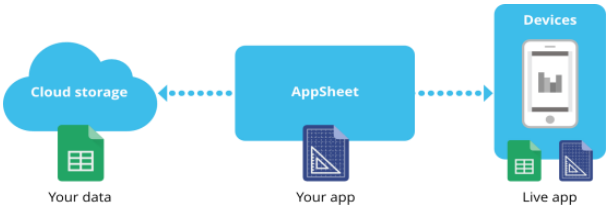


Figure 1: Design Archtecture

Source: Google AppSheet

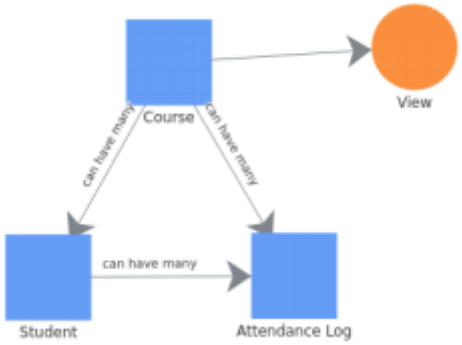


Figure 2: Graph View Presentation

Source: AppSheet

4.2 BACKEND COMPONENT

The backend component used was a google sheet called Coure Manager that has three sheets namely Attendance log, students and course.

Table 1:Google Sheet Tables

SHEET	COLUMN NAMES	SHEET	COLUMN NAMES	SHEET	COLUMN NAMES
Attendance Log	ID	Students	First Name	Course	Course
	Email		Last Name		Description
	Course		Headshot		Image
	Date		Email		
	Time		Course		

Table 1 shows all the sheets for data recording that act as the backend for the application

4.3 FRONTEND COMPONENT

The frontend is the display side of the application where users interact with the app. This works frontend is based on AppSheet. By a single click on the backend interface, an app is created based on the table from Table 1.

Table 2: Frontend Application View

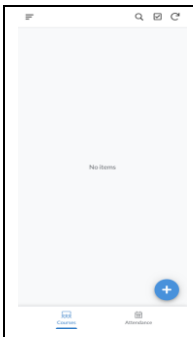
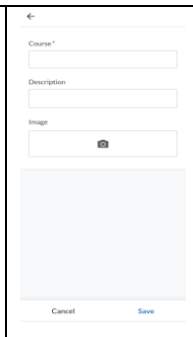
		
Initial screen	Course Creation	Course

Table 2 shows the app after initializing it, course creation and the course appearance. The app has several controls for adding a course and editing the course details.The app allows the instructor to register a student and assign him/her a course.

Table 3:Students Registration Views

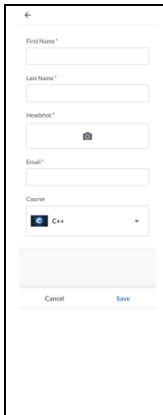
		
Student Registration	Students Profile	Course with students

Table 3 shows students registration and how they appear on the course page and their respective profile.

5. CONCLUSION

In conclusion, this Course Attendance Monitorng Application (CAMA) has been developed successfully to assist course managers in registraton and monitoring of students perticipation in the course. This application is developed with minimal cost which uses google spreadsheet as the main database (backend) storage while Appsheet as application development platform(frontend). This application could be accessed by all the academic staff and making the course manager own their course. For further applications, the course management application could be used by any training institution for the same purpse.

6. REFERENCES

AppSheet. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=AppSheet&oldid=1171803914

Mahboob, T., Irfan, S., & Karamat, A. (2016). A machine learning approach for student assessment in E-learning using Quinlan's C4.5, Naive Bayes and Random Forest algorithms. *2016 19th International Multi-Topic Conference (INMIC)*, 1–8. https://doi.org/10.1109/INMIC.2016.7840094

Martinez, E., & Cisterna, D. (2023). *Using Low-Code and Artificial Intelligence to Support Continuous Improvement in the Construction Industry*. 197–207. https://doi.org/10.24928/2023/0236

Petrović, N., Roblek, V., Radenković, M., & Nejković, V. (2020). *Approach to Rapid Development of Data-Driven Applications for Smart Cities using AppSheet and Apps Script*.

Sitaviana, S., & Indrahti, S. (2022). Optimizing The Performance of Archivist Human Resources Through Appsheet at The Ponorogo Regency Library And Archives Office. *International Journal*, *6*(4).

Sunaryono, D., Siswantoro, J., & Anggoro, R. (2021). An android based course attendance system using face recognition. *Journal of King Saud University - Computer and Information Sciences*, *33*(3), 304–312. https://doi.org/10.1016/j.jksuci.2019.01.006

Wijesekara, D. S., Peiris, P. L. S., Fernando, D. S., Palliyaguru, T. D. N., & Fonseka, W. A. D. N. (2020). Developing an electronic record keeping system at a paediatric clinic in Colombo South Teaching Hospital, Sri Lanka. *Sri Lanka Journal of ChildHealth*,*49*(2),116.https://doi.org/10.4038/sljch.v49i2.8958

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 05 - 11, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1002

Interpolation Fusion Strategy with LSTM for Tax Data Prediction: An Application

Daifang Gou
College of Communication Engineering,
Chengdu University of Information Technology,
State Administration of Taxation Yibin Sanjiang New District Taxation Bureau,
Chengdu, China

Chengyu Hou
College of Communication Engineering,
Chengdu University of Information Technology,
Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China,
Chengdu, China

Linsong Xiao
College of Communication Engineering,
Chengdu University of Information Technology,
Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China,
Chengdu, China

Chang Liu
Chengdu Civil Affairs Bureau Office,
Chengdu, China

Hantao Liu
Education Informationization and Big Data Center, Education Department of Sichuan Province,
Chengdu, China

Wenzao Li
College of Communication Engineering,
Chengdu University of Information Technology,
Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China,
Chengdu, China

Abstract: Traditional data analysis methods have limitations that result in many valuable tax insights being overlooked in routine tax data processing. However, with machine learning techniques, it is feasible to uncover insights that traditional methods fail to capture. Establishing a correlation between historical and future tax data has been a challenging topic, with no effective methods for resolution. Although various models, based on linear and non-linear data, exist for forecasting, they often produce significant errors when applied to tax data predictions.

This paper addresses the non-linear characteristics and volatility of tax data, proposing an enhanced Long Short-Term Memory (LSTM) model. In contrast to the conventional LSTM model, this improved model boasts a higher prediction accuracy. The enhancement involves interpolating the input data and fusing the interpolated data back into the original dataset, aiming to augment the accuracy of the output. In our experimental phase, genuine tax data was used for forecasting, and the superiority of the enhanced LSTM model over the traditional one was visually demonstrated through charts.

Upon predicting tax data for two companies and comparing the outcomes to actual scenarios, it was observed that the proposed enhanced LSTM model significantly reduced the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) by 13.65% and 14.49%, respectively, compared to the traditional LSTM model. This indicates the distinct advantage of the improved LSTM model in enhancing the accuracy of tax data predictions.

Keywords: data analysis methods, tax data, LSTM, accuracy,data predictions.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 05 - 11, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1002

1. INTRODUCTION

1.1 Research Background

Utilizing the LSTM model to analyze tax data introduces an approach harnessing the power of Long Short-Term Memory neural networks to process and interpret data across diverse scenarios [1]. In today's society, the handling and analysis of tax data have become increasingly critical [2]. Tax data, brimming with valuable insights, holds significance for governmental departments, enterprises, and research institutions. However, traditional methods of analyzing tax data present certain limitations, preventing many important tax insights from being fully leveraged. Concurrently, the rise of contemporary technologies like machine learning and deep learning offers new methodologies for a better understanding and analysis of tax data.

Compared to traditional data analysis methods, the LSTM model showcases notable advantages in predicting tax data. Representing a hallmark of deep learning, the LSTM model, equipped with superior sequential modeling capabilities, effectively captures temporal dependencies in tax data, aiding precise future tax trend and variation predictions [3]. Being a nonlinear model, the LSTM contrasts traditional linear models by aptly handling complex nonlinear relations in tax data, thereby enhancing prediction accuracy and reliability [4]. Given the massive and diverse nature of tax data, the LSTM model, with its parallel computational abilities and versatility across data types, efficiently manages large datasets, facilitating comprehensive tax data analysis. Furthermore, its capability to seize and utilize long-term data dependencies is pivotal for analyzing prolonged trends and cyclical shifts in tax data. The real-time and adaptive features of the LSTM model enable on-the-fly parameter updates to navigate the evolving tax environment and influx of new data, proving to be highly adaptable in practical applications.

1.2 Existing Challenges

Traditional tax data analysis techniques are riddled with significant limitations, encompassing data complexity, handling nonlinear relations, capturing long-term trends, and real-time responsiveness. Tax data, inherently complex, consists of a plethora of data types and features. Conventional methods often falter in effectively addressing this data diversity, resulting in information loss or diminished analysis accuracy. Tax data embodies intricate nonlinear relations, potentially influenced by multifaceted factors like economic fluctuations and policy shifts. Classical linear models fall short in grasping these sophisticated nonlinear dynamics, leading to errors in predictions and analyses. Additionally, capturing long-term trends and cyclical changes, essential in tax data, remains a challenge for traditional techniques. Real-time analysis stands as one of the linchpins in tax data interpretation; with a continuous stream of new data, the analytical approach necessitates real-time adaptability to maintain accuracy and relevance [5].

1.3 Proposed Solution

To address the aforementioned challenges, this paper presents an enhanced LSTM network model tailored for tax data analysis. Considering the nonlinear attributes and volatility of tax data, this refined model, compared to its traditional LSTM counterpart, boasts heightened prediction accuracy. The enhancement involves interpolating the model's input data and integrating the post-interpolation data back into the primary dataset, aiming to elevate output data accuracy. Through this methodology, the complex nonlinear relationships within tax data can be better understood, facilitating more precise future tax trend predictions. With these advancements, tax data can be analyzed more effectively, yielding accurate forecasts and profound insights.

1.4 Paper Structure

This paper proposes an improved LSTM model, demonstrating superior accuracy in real-world tax data prediction compared to the conventional LSTM model. The paper's contributions are as follows:

(1) We employ the traditional LSTM model to forecast based on real tax data, presenting the respective predictive results visually in Chapter 4. Comparative results of predictions against actuals, as well as evaluation metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), are discussed [6,7].

(2) Through interpolation of input data and integrating this interpolated dataset as a segment of the primary input, an advanced LSTM model is derived and applied for predictions. Final comparative results, along with evaluation metrics MAE and RMSE, are outlined, juxtaposing the accuracy with that of the conventional LSTM model.

(3) Conclusively, performance evaluations of both LSTM models are conducted. Chapter 6 culminates with an in-depth summary of the characteristics of both LSTM models, emphasizing the merits of the proposed enhanced LSTM model and its quantified advantages.

2. RELATED WORK

In this chapter, we mainly introduce traditional data analysis and forecasting methods and delve into the application of machine learning in tax data prediction. We highlight the current challenges and explore the latest methods to address these challenges.

2.1 Traditional Data Analysis Methods

In the realm of traditional data analysis, the processing of tax data often relies on conventional methods from fields such as statistics and economics. These methods encompass time series analysis, regression analysis, and hypothesis testing [8,9,10]. Time series analysis is extensively utilized to discern temporal trends and cyclical variations in tax data, yet it generally struggles with handling intricate nonlinear relationships. Regression analysis is frequently employed to investigate causal relationships in tax data, but its capability to model interactions among multiple variables and nonlinear relationships is limited. Hypothesis testing methods are used to verify assumptions about tax data, but they may overlook the intricacies and uncertainties inherent in the data.

2.2 Application of Machine Learning in Tax Data Prediction

With the advent of machine learning and deep learning, an increasing number of studies have started to explore their applications in tax data analysis. Machine learning models, such as decision trees, random forests, and support vector machines, have been utilized for the classification and prediction of tax data [11,12,13]. They serve purposes like classification, regression, and forecasting in the context of tax data. Compared to traditional methods, machine learning models are adept at capturing complex relationships within

the data, thereby enhancing the accuracy and reliability of tax predictions. However, they still grapple with dependencies on data feature engineering and challenges in handling time series data.

3. LSTM MODEL AND MODELING BASED ON TAX DATA

In the previous chapter, we discussed traditional data analysis methods and the application of machine learning in tax data prediction. In this chapter, we delve deeper into the LSTM model and its modeling based on tax data, including both traditional LSTM models and modified versions. We will also introduce methods of numerical interpolation processing and discuss how to train and forecast using these models to better comprehend how to utilize them to enhance the prediction accuracy and reliability of tax data.

3.1 LSTM Model Based on Tax Data

In the context of tax data, time series data can be visualized as sequences with continuous time steps, where each step encompasses an observation, such as tax revenue or economic indicators. In this section, we explore the traditional LSTM model tailored for tax data, which is a variant of Recurrent Neural Networks (RNN) and is widely used for sequence data modeling [14]. The LSTM model aims to address challenges faced by conventional RNNs when dealing with long sequence data and long-term dependencies. To counteract the shortcomings of RNNs, LSTM introduces a specialized gating mechanism, comprising input gates, forget gates, and output gates, along with internal memory cells [15]. The role of these gates is to control the flow of information, allowing LSTM to capture long-term dependencies more effectively. The forget gate permits the model to selectively forget previous information, while the input and output gates decide when to introduce new information and when to produce outputs. These mechanisms help alleviate the vanishing gradient problem while retaining sensitivity to long-term dependencies.

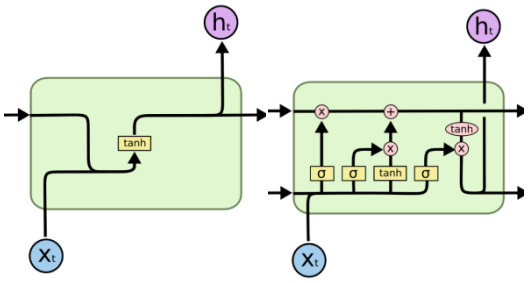


Figure. 1Comparison between LSTM model and traditional RNN model

The architecture of a traditional LSTM (Long Short-Term Memory) model includes an input layer, hidden layers, and an output layer. The input layer typically consists of input nodes representing time steps and other relevant features. The hidden layers are composed of multiple LSTM units, each of which has its own internal state (memory cell) and output gate. The design of these LSTM units allows them to capture long-term dependencies in sequential data, which helps in making more accurate predictions of future trends and changes.

The training process of a traditional LSTM model typically involves steps such as data preparation, model construction, model training, and model prediction. First, time series data is prepared in a format suitable for model input. Then, an LSTM model is constructed with a clear definition of its architecture and parameters. Subsequently, the model is trained using historical tax data by fine-tuning its weights and parameters through the minimization of a loss function to better fit the data. Finally, the trained LSTM model can be used for predicting future tax data. Although traditional LSTM models perform well in handling time series data, they may encounter performance bottlenecks when dealing with complex nonlinear relationships and noisy data. Therefore, the development of improved LSTM models to enhance performance becomes necessary.

3.2 Improved LSTM Model Based on Tax Data

Building upon the traditional LSTM model, we have introduced an enhancement known as data interpolation, which demonstrates improved performance when dealing with tax data [16]. In this enhanced LSTM model, we calculate the average of the past two time steps' data and insert this average as a new data point into the input data. This process is referred to as data interpolation, and its purpose is to smooth the original data, reduce noise in the data, and enhance the training and prediction capabilities of the model. Data interpolation aids in improving the LSTM model's ability to capture the characteristics of the data when dealing with complex nonlinear relationships. By introducing the average value, we can reduce spikes and fluctuations in the data, enabling the model to learn the data patterns more stably. This is particularly useful when handling time series data like tax data, which exhibits periodicity and trends. The improved LSTM model can more accurately capture these trends and periodic changes, thereby enhancing the accuracy of tax data predictions.

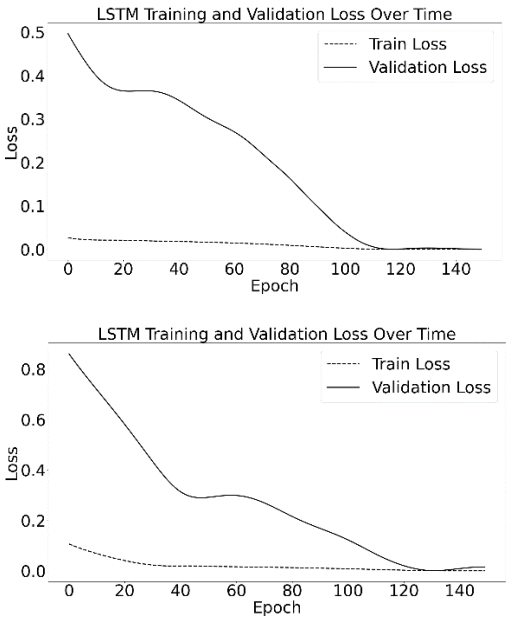


Figure. 2Loss Comparison between LSTM Model and Improved LSTM Model

Figure 2 displays the loss curves during the training process of the traditional LSTM model and the improved LSTM model. It can be observed that the loss trends of both models are similar, but the improved LSTM model exhibits greater stability in some cases. This indicates that the improved model fits the data better during training, reducing the influence of fluctuations and noise. By introducing the enhancement of data interpolation, we can make better use of the information in tax data, gain a deeper understanding of the nonlinear relationships within the data, and enhance the model's performance.

3.3 Model Training and Prediction

In the process of building the improved LSTM model, the training and prediction steps are of critical importance, forming the foundation for the final tax data predictions. Model training is a complex process that begins with data preparation, including cleaning and preprocessing the raw tax time series data to ensure its suitability for input into the LSTM model. The model construction phase involves defining the LSTM architecture and selecting its parameters, including input dimensions, the number of hidden units, and output dimensions. The choice of the loss function and optimizer is also crucial; the appropriate loss function directly affects the model's performance, while the optimizer selection impacts the speed and stability of the training process.

For the input dimension, we selected 1 to represent a single feature (taxable amount), and the hidden dimension (i.e., the number of hidden units) was set to 50. We used Mean Squared Error (MSE) as the loss function and the Adaptive Moment Estimation (Adam) optimizer to optimize the model's weights and parameters. Using the trained LSTM model, we performed sliding window predictions on the entire dataset to obtain future predictions.

Regarding the prediction results, we conducted experimental validation using real tax data from XX city for the years 2019 to 2020. We used the taxable amount for the corresponding months in these two years to predict the taxable amount for each month in 2021. We performed predictions using both the LSTM model and the improved LSTM model, and we visualized the differences between the two models using line graphs and bar charts for Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The improvement in performance was quantified by the difference in MAE and RMSE values.

4. SIMULATION AND DISCUSSION

In this chapter, we will provide a detailed overview of the experimental setup, present the results, and discuss the implications of the experimental findings. Firstly, we will outline the experimental settings, including data sources and the specific objectives of the prediction task. Subsequently, we will showcase the experimental results and quantitatively evaluate the performance improvement of the enhanced LSTM model using metrics such as MAE and RMSE.

4.1 Experimental Setup

The experiments in this study are based on real tax data sourced from XX city's tax records. The objective of the experiments is to use historical tax data (data from corresponding months in 2019 to 2020) to predict future tax data (taxable amounts for each month in 2021). In the experiments, we employed two different LSTM models: the traditional LSTM model and the improved LSTM model, to compare their performance.

The input dimension for the experiments was set to 1 to represent a single feature (taxable amount), and the hidden dimension (i.e., the number of hidden units) was set to 50. The loss function used was MSE, and the Adam algorithm was selected to optimize the model's weights and parameters. For the traditional LSTM model, the input consisted of data from corresponding months in 2019 to 2020 (e.g., taxable amounts for January 2019 and 2020 to predict data for January 2021). For the improved LSTM model, the input dimension, hidden dimension, loss function, and optimization algorithm remained consistent, but we introduced an additional set of interpolated data into the input, namely, the average of data from corresponding months in 2019 to 2020, resulting in 12 new data points. Consequently, we increased the number of input data points from 2 to 3, with the aim of smoothing the original data and reducing spikes and fluctuations to improve prediction accuracy.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 05 - 11, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1002

4.2 Experimental Results

This study conducted experiments using real tax data to compare the performance of the traditional LSTM model with an enhanced LSTM model in tax data prediction. By analyzing the predictive outcomes of both models, several key findings were obtained. Firstly, we employed the traditional LSTM model to predict the tax data for corresponding months from 2019 to 2020, followed by a comparison with the actual taxable amounts for each month in 2021. We also performed the same prediction task using the improved LSTM model. The following are the main findings from the experimental results:

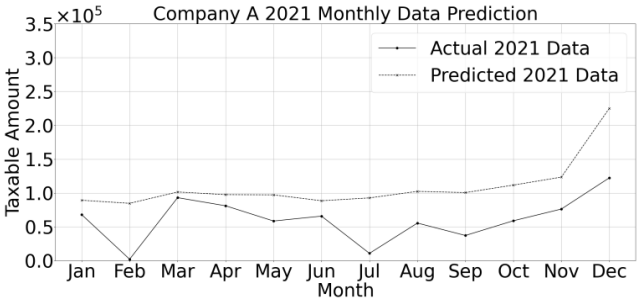


Figure. 3Comparison between LSTM Model Predictions and Actual Data for CompanyA

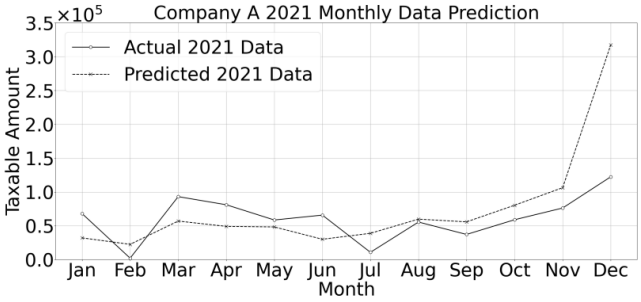


Figure. 4Comparison between Enhanced LSTM Model Predictions and Actual Data for Company A

For Company A, in Figure 3, we observe a comparison between the predictions of the traditional LSTM model and the actual data. It is evident that the general trend is similar, but there is a significant disparity in numerical values. In Figure 4, the enhanced LSTM model's predictions are compared to the actual data, showing an improvement in addressing the issue of substantial numerical disparities.

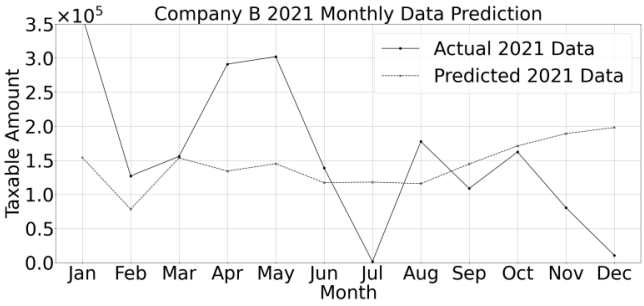


Figure. 5Comparison between LSTM Model Predictions and Actual Data for Company B

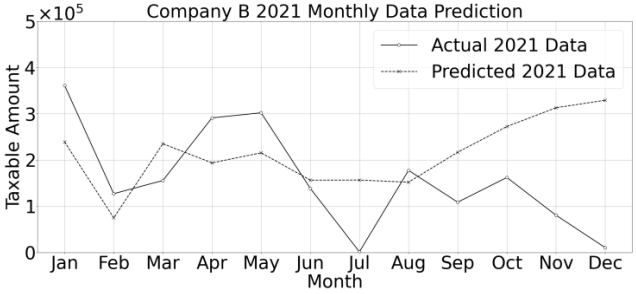


Figure. 6Comparison between LSTM Model Predictions and Actual Data for Company B

The above figures represent the forecast results based on tax data for Company B. In Figure 5, we can observe the comparison between the predictions of the traditional LSTM model and the actual data, especially noting a significant numerical disparity in the later months. Figure 6 shows the comparison between the predictions of the enhanced LSTM model and the actual data, demonstrating a relative improvement in addressing the issue of substantial numerical disparities between predicted and actual tax data.

4.3 Experimental Evaluation

In this section, we will present the performance evaluation results of our proposed enhanced LSTM model for tax revenue prediction. To assess the effectiveness of the model, we employed two commonly used performance evaluation metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics provide valuable insights into the accuracy and precision of the model's predictions.

MAE is a metric that measures the average magnitude of errors between model predictions and actual observations. It is calculated by averaging the absolute differences between the predicted values and the actual values for each data point. A lower MAE indicates that the model's predictions are closer to the actual observed values, suggesting that the model better captures underlying patterns in tax revenue data. The formula(1) for MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_{actual} - Y_{predicted}| \quad (1)$$

Where, n represents the total number of data points. Y_{actual} represents the actual tax revenue value. $Y_{predicted}$ represents the predicted tax revenue value.

RMSE is another widely used metric that takes into account the squared errors between predicted values and actual values, providing a measure of the magnitude of these errors. RMSE is particularly sensitive to large errors because it involves the square of the differences, penalizing significant deviations from actual values. The calculation formula(2) for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |Y_{actual} - Y_{predicted}|} \quad (2)$$

Where, n represents the total number of data points. Y_{actual} represents the actual tax revenue value. $Y_{predicted}$ represents the predicted tax revenue value. Using MAE and RMSE as performance evaluation metrics allows for a reasonably accurate quantification of performance variations.

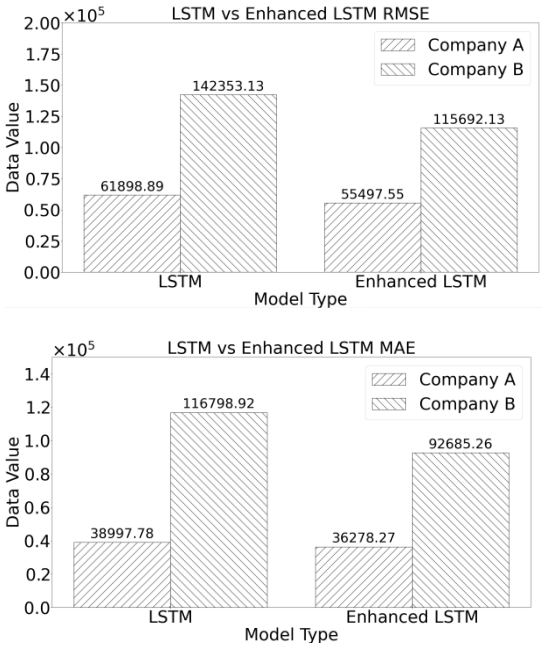


Figure. 7MAE and RMSE Metrics for Predictive Results

From Figure 7, we can observe the numerical comparison of performance evaluation metrics for the two companies under the two LSTM models. It is evident that under the improved LSTM model predictions, both MAE and RMSE show a decrease. In the case of Company B, the reduction in both metrics is more pronounced, and lower MAE and RMSE values indicate a better fit between predicted and actual data. Therefore, the enhanced LSTM model demonstrates superior predictive performance compared to the traditional LSTM model.

4.4 Advantages of the Enhanced LSTM Model

Based on the results and discussions from this experiment, we have identified several key advantages of the enhanced LSTM model compared to the traditional LSTM model when dealing with tax data. These improvements are primarily reflected in the following aspects:

Data Interpolation: The enhanced LSTM model incorporates data interpolation, which smoothes the raw data and reduces noise in the data, thereby enhancing the model's training and prediction capabilities. Data interpolation helps the improved model better capture the features of the data, especially when dealing with complex nonlinear relationships.

Better Fitting: By introducing data interpolation, the enhanced LSTM model can better fit the complex nonlinear relationships within tax data, leading to more accurate predictions of future tax trends. The improved model achieves a better fit to the data during the training process, reducing the impact of fluctuations and noise.

Performance Enhancement: Experimental results demonstrate a performance improvement of the enhanced LSTM model compared to the traditional LSTM model in tax data prediction. The average reductions in MAE and RMSE are 13.65% and 14.49%, respectively. This indicates that the improved model more accurately captures the characteristics and trends in tax data.

These advantages highlight the effectiveness of the enhanced LSTM model in enhancing predictive accuracy and handling complex tax data, making it a valuable tool for tax revenue forecasting.

5. CONCLUSION

In this study, we explored the application of traditional data analysis methods and machine learning in tax data analysis and prediction. We introduced traditional data analysis methods, including time series analysis, regression analysis, and hypothesis testing, along with their limitations when dealing with tax data. Subsequently, we delved into machine learning models, particularly the LSTM (Long Short-Term Memory) model, and its application in tax data prediction. We proposed an enhanced LSTM model that incorporates data interpolation to smooth raw data, reduce noise, and improve the model's training and prediction capabilities.

The experimental results demonstrate that the enhanced LSTM model outperforms the traditional LSTM model in tax data prediction tasks. Through data interpolation, the model better captures the nonlinear relationships within tax data, reduces prediction errors, and enhances predictive accuracy. The average reductions in Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are 13.65% and 14.49%, respectively, indicating the superiority of the improved model in tax revenue prediction.

6. ACKNOWLEDGEMENTS

We would like to express our heartfelt gratitude to all those who have contributed to this research. Firstly, we would like to thank our supervisor for his/her valuable guidance and insightful comments throughout the entire research process.

We also extend our sincere appreciation to all the participants who have generously devoted their time and effort to provide us with valuable data and insights. Finally, we would like to thank our families and friends for their unwavering support and encouragement. This research would not have been possible without their support. Also, we thank the lab funding from “Meteorological Information and Signal Processing Key Laboratory of Sichuan Higher Education Institutes of Chengdu University of Information Technology”; “Integrated Computing and Chip Security Sichuan Collaborative Innovation Center of Chengdu University of Information Technology”. The Open Project of National Intelligent Society Governance Testing Area (NO. ZNZL2023A04). the fund of the Scientific and Technological Activities for Overseas Students of Sichuan Province 2022(30) and funds from the Sichuan Provincial Department of Human Resources and Social Welfare” Researches on Key Issues of Edge Computing Server Deployment and Computing task Offloading”. Thank you for the funding, named "Research on the Construction and Management of Smart Civil Affairs Shared Data System" of Chengdu Civil Affairs Bureau. And Sichuan Science and Technology Program, Soft Science Project (No.2022JDR0076).

7. REFERENCES

1. Shu., Xiangbo, Zhang,.Liyan, Sun., Yunlian,Tang,,&Jinhui.(2021).Host-Parasite: Graph LSTM-in-LSTM for Group Activity Recognition.IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,32(2),663-674.
2. Blanchet, Thomas, Flores,.Ignacio,Morgan., & Marc.(2022).The weight of the rich: improving surveys using tax data.JOURNAL OF ECONOMIC INEQUALITY,20(1),119-150.
3. Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*, 48-62.
4. Vaupel, Y., Hamacher, N. C., Caspari, A., Mhamdi, A., Kevrekidis, I. G., &Mitsos, A. (2020). Accelerating nonlinear model predictive control through machine learning. *Journal of process control*, *92*, 261-270.
5. Szabó, B., &Babuška, I. (2021). Finite Element Analysis: Method, Verification and Validation.
6. Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C. H. (2020). On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Processing Letters*, *27*, 1485-1489.
7. Ćalasan, M., Aleem, S. H. A., &Zobaa, A. F. (2020). On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function. *Energy conversion and management*, *210*, 112716.
8. Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., ... & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research*, *50*(W1), W276-W279.
9. Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39-42.
10. Seiffert, D. J., McCarthy Veach, P., LeRoy, B., Guan, W., &Zierhut, H. (2019). Beyond medical actionability: Public perceptions of important actions in response to hypothetical genetic testing results. *Journal of Genetic Counseling*, *28*(2), 355-366.
11. Charbuty, B., &Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28.
12. Probst, P., Wright, M. N., &Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, *9*(3), e1301.
13. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189-215.
14. Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, *31*(7), 1235-1270.
15. Yang, Y., Li, L., Lin, R. B., Ye, Y., Yao, Z., Yang, L., ... & Chen, B. (2021). Ethylene/ethane separation in a stable hydrogen-bonded organic framework through a gating mechanism. *Nature Chemistry*, *13*(10), 933-939.
16. Wang, B., Zhang, N., Lu, W., & Wang, J. (2019). Deep-learning-based seismic data interpolation: A preliminary result. *Geophysics*, *84*(1), V11-V20.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 12 – 19, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1003

Sentiment Analysis Model for Farmers-Herders Crisis

Ukeyima, Dooshima Joyce
Department of Mathematics and Computer Science, Benue State University, Makurdi, Nigeria

Obilikwu, Patrick
Department of Mathematics and Computer Science, Benue State University, Makurdi, Nigeria

Kile, Awuna Samuel
Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria

Abstract
The famer-herder crisis over scarce grazing land resources has over the years resulted to the loss of several lives and properties as well as reduction in food production in Benue State. Social media plays a key role in conflict escalation or de-escalation because they are information carriers. In this study, Sentiment analysis is used to analyse opinions of social media comments and posts about farmers-herders crisis expressed by Twitter users to ascertain the positive, negative and neutral tweets. Tweets were extracted using Python library called Snscrape, model was designed and processing was done using the Natural Language Tool Kit (NLTK) to categorize the data into positive, negative and neutral. The naïve Bayes and random forest were used for the analysis and evaluation was done using Receiver Operating Characteristics (ROC), F1 and Accuracy. The values for ROC, F1 and Accuracy for Random Forest and Naïve Bayes algorithms are 0.73, 0.61 and 0.90, and 0.70, 0.73 and 0.60 respectively. It was observed that Random Forest performed better than Naïve Bayes, and as such, can be applied towards assisting security agencies and government policies on the crisis management.

Keywords: Classification, crisis, positive, negative, neutral, tweets, sentiment analysis.

1.0 INTRODUCTION

With a population of 4,219,244 residents as of the 2006 Census, Benue State is a Nigerian state that is situated in the North Central Zone. 1976 saw the creation of the state (Commission, 2006). Its entire size is roughly 33,955 square kilometers on land. Since the majority of Benue State's indigenous population (the Tiv, Idoma, Igede, etc.) are farmers, subsistence farming serves as their primary means of subsistence. The majority of Benue State's large arable area is often used for crop cultivation, with most common crops being rice, yam, and cassava. One of the states in the nation's guinea savannah vegetation zone, Benue State is known for its long grasses and lack of swamps, making it ideal for cattle grazing. Due to these, the number of Fulani nomadic livestock breeders entering the State has increased steadily. The confrontations between farmers and herds in the Benue area have been more violent recently (Eti & Akpu, 2021). Ugwumba (2008) claims that throughout time, struggles for limited land resources and grazing land have led to persistent violent conflicts that are expanding in terms of frequency, intensity, and geographic reach.

According to (Eti & Akpu, 2021), between September 2017 and June 2018, the Nigerian herders and farmers conflict resulted in over 1,500 deaths and over 300,000 displaced people who are currently living as internally displaced persons (IDPs) in various camps, settlements, and communities. These conflicts have claimed several lives and properties in the state. The issues of food insecurity from the standpoint of food production has been made worse by the violence committed by herdsmen and native farmers in

Benue State, which has had a severe impact on national food production (Alao et al, 2019).

Babale and Nasisdi (2019) believe that during conflict, knowledge plays a crucial role in either increasing or decreasing tension. The main channel via which attitudes (i.e., biased remarks, posts, or comments) have been eradicating the current generation is social media. Because social media usage in Benue State is so unrestricted, its effects have been felt widely. This explains the recent spate of overly emotive remarks about farmers' and herders' rampage. Even while these crises have serious repercussions for the herders, indigenous farmers, and the country as a whole, occasionally they are exaggeratedly reported on social media platforms like Facebook, Twitter, Instagram, WhatsApp, Messenger, and so forth. In contrast to traditional media, where editors strictly monitor and uphold professional and ethical standards, news content is subject to internal and external scrutiny. Avoiding statements and messages that could jeopardize national unity, peace, and development is the goal here (Annual, 2018). The majority of the works that have already been done either center on rules (Krause & Grassegges, 2010) or manual feature extraction (Gitari et al, 2015). Rule-based techniques usually rely on a precompiled list or dictionary of clues and do not require learning (Haralambous & Lenca, 2014). Mehdad & Tetreault (2016) employed support vector machines (sum) to identify sentiment analysis after extracting the text's n-gram,

character level, and sentiment features. Artificial features, on the other hand, are limited to reflecting the surface features of text and are unable to comprehend content derived from deep semantic features.

In general, one is persuaded that the great freedom offered by social media is, in fact, mysterious by closely examining the comments. This makes it possible for disparaging remarks to be heard, which leads to the emergence and spread of outright hatred and denigration. Youths in Benue State have been reacting to herders' issues on social media, claiming that the Fulani herdsmen are staking claims to the Benue valley.

This research project proposes a data mining technique called sentiment analysis to filter out biased statements that could exacerbate the conflicts between farmers and herders in Benue State. The goal is to then provide commendable solutions to the conflicts raging the state, given the detrimental effects of fake and biased news on social media.

This aids in monitoring the polarity or duality of such data to prevent tension from unnecessarily increasing due to emotional remarks made in relation to such data. Every social media post about the problems facing Benue State's farmers and herders is scrutinized for subjectivity. In light of these worries, this study aims to characterize and contextualize the plethora of remarks and postings made on social media platforms related to the conflicts facing farmers and herders in Benue State. "Twitter" is one of the most popular social networking sites in Benue. More than 330 million people use Twitter every month (Statista, 2019). While some other social media platforms require both parties' approval in order to connect, Twitter allows users to compose brief messages within 280 characters, known as tweets (Statista, 2019). Without requiring permission, Twitter users can follow submissions—also known as tweets—from other users. It is one of the reasons Twitter is the hub for information sharing; it serves as a forum for interacting with the public and other officials as well as promoting political ideas. examining Twitter data to determine popular opinion regarding the problems facing farmers and herders. It is possible to accurately assess the dilemma facing farmers and herders, make improvements, and stop the issue from getting worse.

The major security threat Benue State now face is the perennial conflict with Fulani herders. A number of cases of killings, farmlands destruction, raping, and displaced indigenes have been reported on different social networking sites at different times. The cause of this current bout of clashes remains an issue of debate among Nigeria watchers. The conflict is becoming fierce and increasingly wide spread in Nigeria and the more social media places emphasis on an issue or event, the more people see such issues as important hence sentiment analysis focused on discovering the influence of individual feelings on social media and understanding of social realities in general. This research on farmers and herder's crisis seeks to peruse the numerous comments and posts made via social media platforms. It identifies each statement as positive, negative and neutral using data mining technique called sentiment analysis so as government can provide decision on the conflict.

The aim of the study is to develop a sentiment analysis or opinion mining model to analyse social media comment and posts about farmers/herder crises in Benue State.

The objectives of the study include:

To extract and gather social media comments and post about farmer/herder crises in Benue

To develop a model to analyze the sentiment

To perform sentiment classification on the tweets into positive, negative and neutral.

Evaluate the performance of the proposed system

This study is significant because the government can use the knowledge gotten from the findings of the research to understand the feelings of the populace about farmers/ herders' crises and how it affects them within the state. This can assist governmental policy makers in formulating policies and encourage peaceful co-existence between herdsmen and indigenous farmers.

2.0 LITERATURE REVIEW

Sentiment analysis, also known as opinion mining, is the process of identifying and categorizing subjective opinions in source materials (such as documents or sentences) by using natural language processing, computational linguistics, and text analytics, according to Luo & Xu (2019). Sentiment analysis generally seeks to ascertain the writer's attitude toward a particular subject or the document's overall contextual polarity. The author's attitude can be interpreted as their assessment or judgment, their affective state (i.e., how they are feeling when writing), or their intended emotional communication (i.e., the emotional impact they want to leave on the reader).

Guyal (2020) defines sentiment analysis as the computational analysis and identification of views and judgments within a text. Based on sentiment analysis, you may determine whether a text is favorable, negative, or neutral.

Four categories were created by Kah and Zeroual (2022) to group sentiment analysis into:

Sentiment Analysis using Fine Grins: The precision of polarity is determined using sentiment analysis. The following polarity categories can be used for sentiment analysis: extremely positive, positive, neutral, negative, or very negative. Review and rating analysis benefits from fine-grained sentiment analysis. On a rating system of 1 to 5, for instance, 1 would be seen as extremely negative and 5 as very favorable.

Aspect-Based Sentiment Analysis: Aspect-based analysis goes further than fine-grained analysis in identifying the general polarity of customer evaluations. It aids in identifying the specific topics that individuals are discussing. A consumer review for a smartphone might say, for instance, that "the camera struggles in artificial lighting conditions." It can be ascertained through aspect-based analysis that the reviewer made a comment regarding a "negative" element of the camera.

Emotional Detection Sentiment Analysis: Emotion detection aids in the identification of emotions, as the name implies. Anger, despair, joy, frustration, dread, anxiety, panic, etc. are examples of this. Lexicons, a list of words that represent specific emotions, are commonly used by emotion detection systems. Robust machine learning (ML) algorithms are also used by some advanced classifiers. Take the claim that "This product is about to kill me," for instance. could be taken to mean having panic and fear.

Intent Analysis: Companies can save time, money, and effort by accurately determining the intent of their customers. Businesses frequently wind up pursuing customers who have no intention of making a purchase anytime soon. Precise purpose analysis can overcome this obstacle. The purpose analysis aids in determining the consumer's intent, including whether they plan to buy or are only perusing. He can follow the customer and target them with ads if they are inclined to buy.

Verma (2018) distinguished between three main methods for sentiment analysis: hybrid, machine learning, and lexicon-based methods.

Lexicon-Based Approach: Opinion-bearing words are gathered and collected for use in lexicon-based sentiment analysis of text, a data analysis activity that uses opinion words and phrases without any prior information. Opinion lexicon refers to both positive and negative terms as well as opinion expressions. Unlabeled data and lexicon were employed in the lexicon-based technique. Words in the text are assessed according to opinion lexicon in order to ascertain their orientation and, consequently, the text's sentiment. The creation of opinion lexicons is essential to the lexicon-based sentiment analysis method. One of these three methods is typically used to generate opinion lexicons in lexicon-based systems.

Manual Approach: Opinion words are personally gathered according to each person's proficiency with the language and domain knowledge. This procedure takes a lot of time. This method is typically used in conjunction with automated methods to correct errors made by the latter.

Dictionary-Based Approach: Words of opinion with established orientations are gathered from lexicographical sources such as internet dictionaries. To ascertain word feelings, it makes use of opinion lexicon hierarchies, synonyms, and antonyms. Dictionary-based techniques have limitations in recognizing sentiment particular to a context because they lack domain knowledge. WordNet, SentiWordNet, secticNet, sentifull, and other dictionaries might be employed.

Corpus-Based Approach: A corpus-based approach finds and gathers opinion words in a huge corpus by using the syntactic pattern of opinion words and co-occurrence words. The context-specific classification of opinion words in dictionary-based approaches is eliminated by the corpus-based method. On the other hand, dictionary-based methods are more effective. Labeled data was employed in the corpus-based technique.

Machine Learning (ML) Approach: The foundation of these learning strategies is the construction of classifiers from labeled textual post examples. For the domain they are trained in, they function well. There are two categories of machine learning approaches.

Supervised learning: Labeled training papers are used in supervised learning techniques. The foundation of supervised learning is automatic text classification. Utilized is a labeled training set with pre-established categories. A classification model is developed to forecast a document's class by using a pre-established category. Algorithms for supervised learning include Support Vector Machine, Random Forest, Decision Tree, and Naïve Bayes.

3.0 METHODOLOGY

The study adopted the qualitative and analytical method since the nature of data gathered is expressed as opinions and analysed.
Python Snscrape library was used to mine data from twitter between a time period. Data was cleaned/ tokenized.
Sentiment detection was done, Natural Language Processing Tool Kit (NLTK), opinion was examined for subjectivity, polarity and was classified into positive, negative and neutral.
The system design is achieved using Unified Modelling Language (UML).
Supervised learning approach of Naïve Bayes algorithm and Random Forest was employed in evaluating the result in python programming language.
Receiver Operating Curve (ROC) using the True and False Positives of Naïve Bayes is used to evaluate the performance of the result of the Random Forest.

3.2 Analysis of Proposed System

The Model was developed using Python's Natural Language Processing Tool Kit (NLTK) and Naïve Bayes Classifier algorithm and Random Forest were used to analyse the data.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 12 – 19, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1003

NLTK is a leading platform for building python programs to work with human language data. it provides easy to use interfaces to 50 corpora and lexical resources such as WordNet, along with a suite of word processing libraries for classification, tokenization, stemming, tagging, parsing and semantic reasoning. The sentiment analysis model was developed to classify social media (Twitter) comments on the farmers/herders' conflict in Benue State into POSITIVE, NEGATIVE and NEUTRAL on a -1 to 1 scale where -1 represents negative comments, 0 represent neutral and 1 represents positive comments. The input dataset consists of tweets of individuals who expressed their feelings and concerns on the conflict on Twitter. The tweets on the farmers/ herders' conflicts in Benue State were extracted from the Twitter using a python library called Snscrape.

3.2.1 How to Scrape Tweets from a User or text search with Snscrape

Snscrape includes two methods for getting tweets from Twitter: the command line interface (CLI) and a Python Wrapper. The sntwitter. TwitterSearchScaper () is the method that returns an object of tweets from the name of the user we passed into it (example John) or a phrase written. Snscrape does not have limits on numbers of tweets so it will return however many G6

tweets from that user. To help with this, we need to add the enumerate function which will iterate through the object and add a counter so we can access the most recent 100 tweets from the user. Snscrape also allows us to pass in the date from which we want to start the search and the date we want it to end in the sntwitter.TwitterSearchScraper() method.

3.3 Architecture of the proposed Model

The architecture of the model is shown in figure 1 below

Figure 1, the architecture of the proposed model consists of the cloud which represents the social media from which the farmers/ herders' conflicts data (tweets) were extracted; the Python Snscrape API which was used to extract or scrape the dataset from the social media, preprocessing (tokenized) was done the NLTK was used to classify the dataset, Naïve Bayes and Random Forest was used to evaluate the results. The steps in building the model can be represented algorithmically as shown below:

(i) Start
(ii) Crawl social media posts from twitter
(ii) Tokenise or pre-process the dataset
(iii) Detect sentiment post or comments
(iv) Classify the detected sentiment post or comments
(v) Display results
(vi) End

3.3.1 Design of the Model

The design o is achieved using Unified Modeling Language (UML) tools like use-case diagram, sequence and class diagrams. Tools employed are StarUML and Microsoft word.

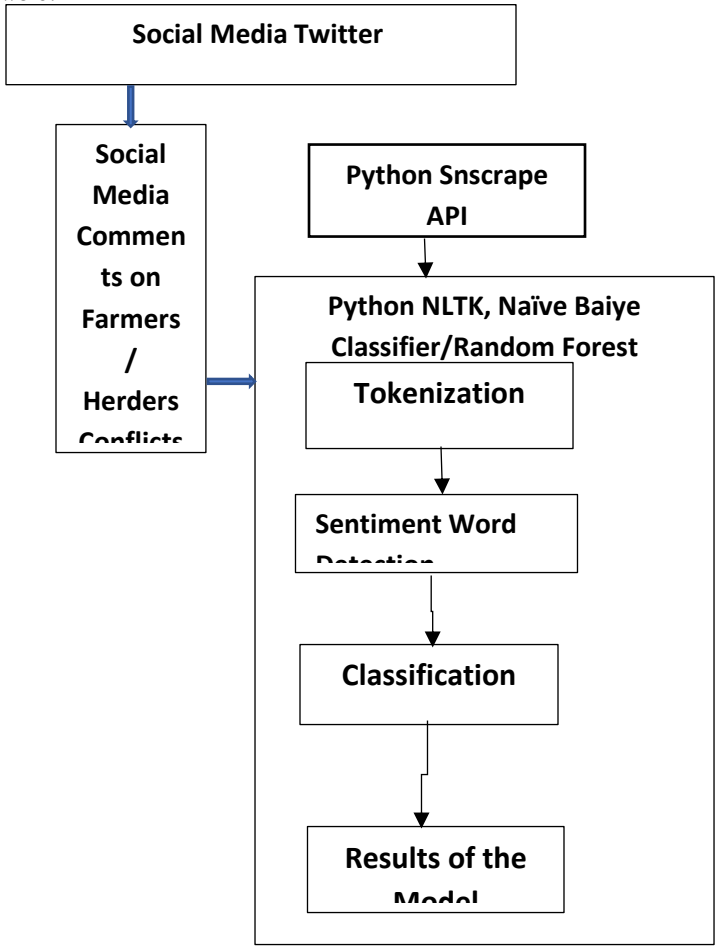


Figure 1: Architecture of the Proposed System

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 12 – 19, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1003

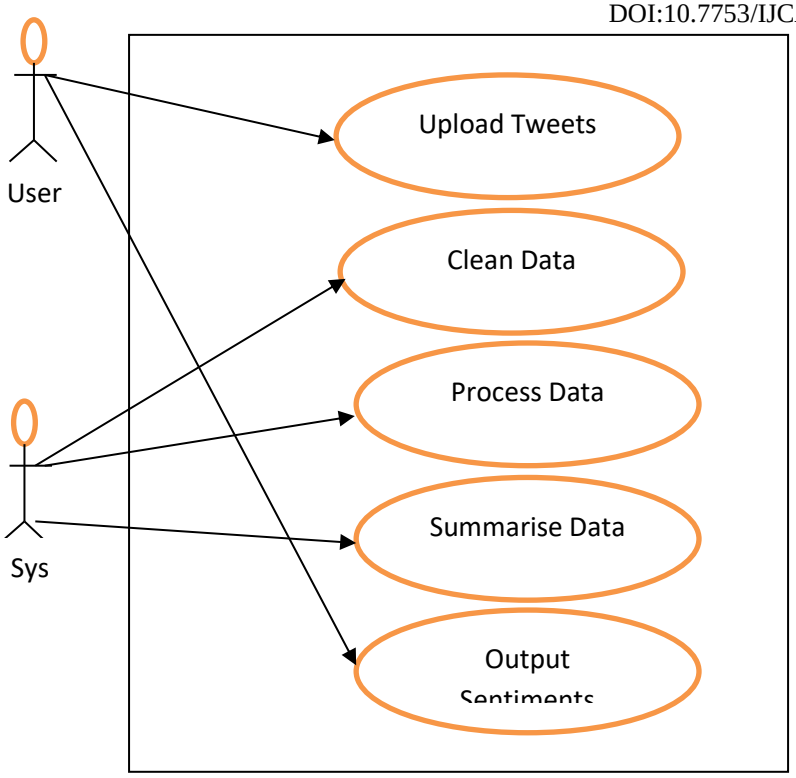


Figure 2: Use Case diagram for the Sentiments Analysis of Farmer-Herder Crisis

The User can upload or provide tweets to web system through the API. He can also see how the sentiments will be displayed, whether they are more of positive or negative, and from there guide the user towards the action that maybe taken on the stated subject. The system can perform functions like data cleaning, preprocessing and processing, summarizing and also displaying the output of the sentiments.

3.4 Sentiment Analysis classification using Natural Language Tool Kit (NLTK)

Steps below summaries how opinion mining can be performed using NLTK.

1. Importing Dataset: This is done using python's Panda library. Data is extracted and saved in a Csv file. Sample of data is shown

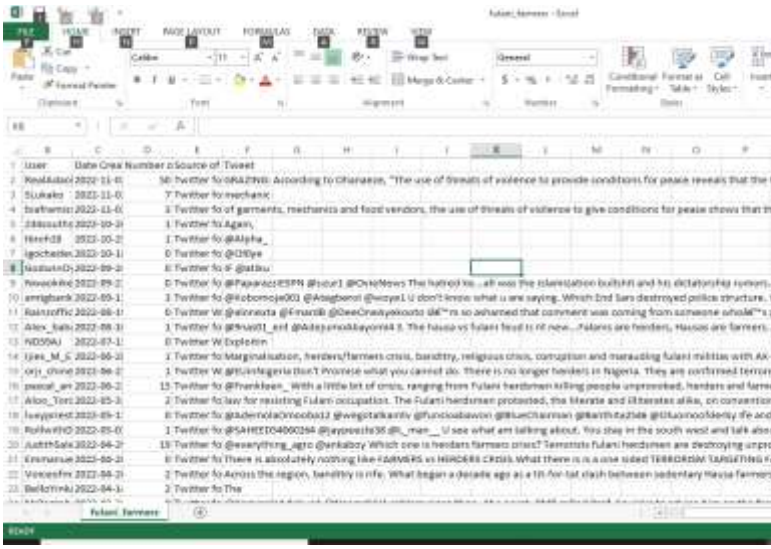


Figure 4: Sample of imported data

2. Data Processing and Visualization: After having access to the data, the data is cleaned, Punctuations and stop words are also cleaned and Classification is done with the help of NLTK library.
3. Prediction: To clarify the performance of our model, we supply different statements and allow our model using predict if its negative, positive or neutral, with -1 for negative, 1 for positive and 0 for neutral.
4. Evaluation: This was done using Naive Bayes and random forest Algorithm. Both Algorithms describe additions in detail through indications of accuracy, precision Receiver Operating Curve (ROC) in python sklearn module.

4.0 SYSTEM IMPLEMENTATION

The implementation of the system is achieved using Python programing language. SnScrape library in the Pycharm IDE is used for the data extraction. Naives Bayes and random forest algorithms are applied in the data processing of the farmer-herder crisis management system. Results gotten are tweets from Twitter about farmer-herder crisis and they are processed to determine their polarities, whether the sentiments are negatively or positively inclined, thereby enabling the decision makers towards making informed decisions.

4.1 System Implementation Algorithm

The algorithm for system is as given below.
Step 1: Collect data about farmer-herder crisis from Twitter using tools, functions and algorithms
Step 2: Process the collected data
Step 3: Analyze the collected data to give you the polarities
Step 4: Visualize the processed data to bring out meaning from the sentiments.

4.2 System Results

Results obtained from the system are presented in this section. Screenshots and graphical illustrations presented in the diagrams below are results of extracted processed tweets about farmer-herder crisis in Benue State. Details of the

processed results are:
Total tweets extracted: 576
Duration of extraction: From 2020 to 2022
Number of positive sentiments: 183
Number of negative sentiments: 226
Number of neutral sentiments: 167
The visualization of the processed extracted data is thus presented in screenshots from the system and evaluation diagrams as shown.

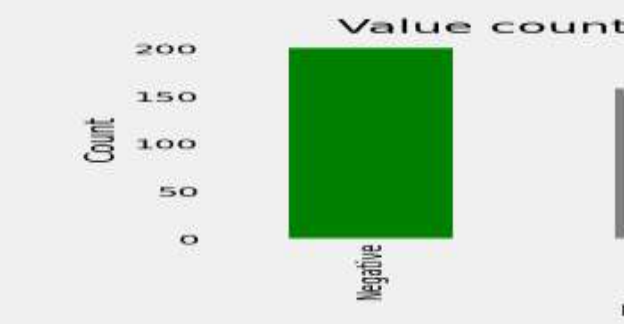


Figure 5: Bar chart Representation of the polarities of Twitter Sentiments of Farmer-Herder Crisis

Figure 5 shows the processed word count representation and the polarities. It can be seen that polarities are represented in bars and distinguished by colors with 'green', 'ash' and 'red' representing the negative, positive and neutral polarities respectively. Extracted and processed words count shows that negative sentiments were more and should affect decision taking accordingly.

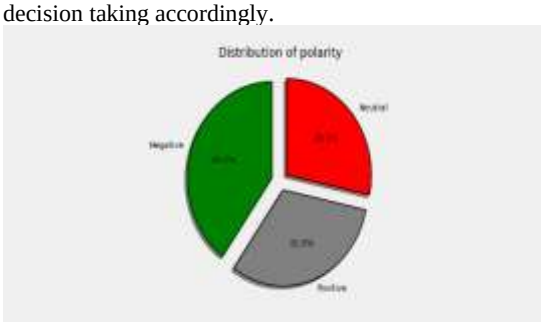


Figure 6: Pie Chart distribution of the Sentiments Analysis

Figure 6 shows the percentage distribution of processed polarities. The negative, positive and neutral polarities have percentage distributions of 40.1%, 31.5% and 28.3% respectively. This indicates that the negative polarity has more representation in the processing of extracted words from Twitter giving credence to the severity of the crisis calling for urgent attention.

Scatter Graph of Subjectivity vs Polarity

It shows output for all tweets, while polarity is concentrated at the centre, subjectivity is mostly spread at the edges, it shows a negative correlation than the positive correlation.

Figure 7: Scatter Graph of Subjectivity versus Polarity showing the analysis of the sentiments, shows the scatter of subjectivity against polarity

Word Cloud

This was created to understand which words have been used in most of the tweets.

Figure 8: Word cloud

Top Words

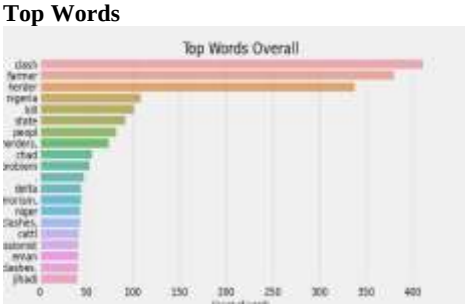


Figure 9: Farmer-Herder Top Words Extracted from Twitter

Figure 9 shows words extracted from Twitter and processed about farmer-herder crisis. It is bar chart and the words are plotted against the total counts. The words 'clash', 'farmer' and 'herder' appears most as can be seen from the word cloud giving an indication that words extracted and processed from Twitter about this subject are correctly extracted.

ROC CURVE

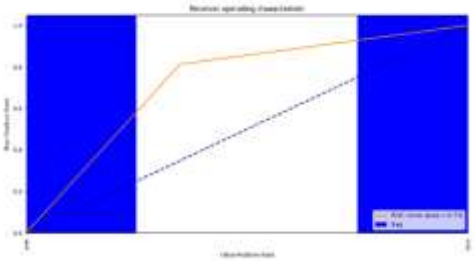


Figure 10: Sentiment Analysis Evaluation of Twitter Farmer-Herder Crisis System

Figure 10 shows the evaluation of the farmer-herder crisis sentiments analysis system. It applied the Receiver Operating Characteristic (ROC) and plotted 'True Positive Rates' against 'False Positive Rates'. The ROC area under the curve was 0.73.

4.3 System Performance Evaluation

The system developed in this study is evaluated using ROC, F1 and accuracy.

The table of performance evaluation of the developed system given below:

Table 4.1: Performance evaluation of the developed system

	ROC	F1	Accuracy
Random Forest	0.73	0.61	0.90
Naïve Bayes	0.70	0.73	0.60

From the table 1.0, the values of ROC and accuracy for the random forest algorithm are higher than the Naïve Bayes Algorithm; the F1 value for random forest is lower than that of naïve bayes. Since the ROC and accuracy of the random forest algorithm is better than that of naïve bayes, it therefore means that the random forest algorithm performed better than the naïve bayes algorithm in this study.

4.4 Discussion of Results

Results obtained for the sentiment analysis system are inform of total tweets extracted, number of positive tweets, number of negative tweets, number of neutral tweets and their representations pictorially. The system was implemented using two algorithm and the results of their performance compared. Total number of tweets extracted were 576 for the duration of 2020 to 2022. Number of positive sentiments were 183, number of negative sentiments, 226 and number of neutral sentiments gotten was 167.

The processing of these tweets was achieved using two algorithms, random forest and naïve bayes. Results of processing were gotten based on receiver operating characteristics (ROC), F1 score and accuracy.

Results from the random forest showed that ROC had a value of 0.73, the F1 score was 0.61 and the accuracy of the algorithm stood at 0.90. The second algorithm, the naïve bayes algorithm shows a ROC of 0.70, F1 score of 0.73 and the accuracy value of 0.60. This shows that the first algorithm, random forest in this case performs better than naïve bayes. Impliedly, random forest should be adopted and used in similar scenarios. The ROC curve shows a tradeoff between sensitivity (TPR) vs 1-specificity (FPR). The area under the curve is 0.73 indicating an accurate test and the value is good and it suggest that, 73% chance that the model predicted accurately. The dashed line in the ROC curve shows random classifier consistency.

The above results representation shows that tweets extracted from Twitter contained key words that can be analyzed to proffer solutions or prevent a dangerous occurrence of the crisis if systems like this are put in place and monitored real time to obtain information and make informed decisions. The system evaluation using ROC plot of 'True Positives Rates' against the 'False Positive Rates' also gives an acceptable of 0.73.

5.1 SUMMARY

The study looked at sentiment analysis and how it can be applied towards tackling serious problems bedeviling human lives. Specifically, farmer-herder crisis in Benue State was looked into. Social media in shaping lives and events and as such, can be considered as a precursor towards many issues around human. Twitter was adopted and applied in this study. Tweets about farmer-herder crisis were extracted and analyses for emotions or opinions as expressed by Twitter users. Naïve Bayes and random forest algorithms were used in evaluating the extracted tweets. Negative polarity tweets surpassed both the positive and neutral polarities. A model for sentiments analysis operations for farmer-herder crisis management was also developed. This clearly indicates that the farmer-herder crisis problem is a serious problem deserving all approaches in solving and preventing the crisis. The evaluation of the system using Receiver Operating Characteristics (ROC), F1 and accuracy using two algorithms, that is Random Forest and Naïve Bsayes give the values 0.73, 0.61 and 0.90, and 0.70, 0.73 and 0.60 respectively, indicating that the random forest algorithm performed better than the naïve bayes algorithm since the ROC and accuracy are higher as compared to the other algorithm, as such, random forest algorithm may be adopted in similar situations to give better results.

Major highlights of the research are:

- The paper proposes an approach that uses Python Natural Language Processing Tool Kit and machine learning algorithms to analyze and evaluate social media (tweets) opinions that are expressed during farmer/herder crisis in Benue State.
- The extracted tweets are processed using Python's Natural Language Processing Tool Kit, and the major categorization is considered as the people's opinion.
- The opinions are categorized into positive, negative and neutral.
- The performance evaluation shows that the random forest algorithm performed better than the naïve bayes algorithm since it returned higher values for ROC and accuracy.

5.2 Conclusion

This study has achieved the development of a system for managing farmer-herder crisis in Benue State. Sentiment analysis is the technical approach used here and opinions of social media users about the crisis especially Twitter were considered. The data gotten was analyzed using machine learning classification algorithms, Naïve Bayes and random forest. It is evident that this system developed can be a guide to decision makers in providing solutions to this hydra-headed

problem. As such, decision makers, researchers and other interested parties are encouraged to make use if the services offered by the system in tackling these problem and other related ones.

5.3 Recommendations

The government can use the finding of the research to understand the feelings of the populace about farmers/herders' crises and how it affects the state. Further research can also be done in this area using other scientific methods, approaches and technologies. Also, other social media handles can be adopted for opinion mining for different scenarios.

6 REFERENCES

[1] Verma, B. and Thakur, R. S. 2018. Sentiment Analysis Using Lexicon and Machine Learning Based Approaches: A Survey. Proceedings of International Conference on Recent Advancement on Computer and Communication, Lecture Notes in Networks and Systems 34, retrieved from https://doi.org/10 on 26-01-2023.

[2] Guyal, K. 2020. Top 4 Types of Sentiment Analysis & Where to Use. Retrieved from https://www.upgrad.com/blog/types-of-sentiment-analysis/ on 26-01-2023

[3] Alao, D. O., Shaibume, B., Ogunwemimo, T., Alao, E. M. and Ogunwemimo, O. 2019. Herdsmen/Native Farmers' Violence in Benue State and Food Security in Nigeria. Mediterranean Journal of Social Science, 10(6).

[4] Babale, A. M. and Nasidi, Q. Y. 2019. Role of Mass Media towards Ending Fulani Herdsmen and Farmers Conflicts in Benue State. KIU Journal of Humanities, 4(1): 15–22.

[5] Commission, N.P. 2006. National Population Statistics, https://en.m.wikipedia.org

[6] Eti, C. E. and Akpu, B. U. C. 2021. Agro – Pastoralist Conflicts in Benue State Nigeria: Hazards and Imperative for Human Rights Protection. International Journal of Management, Social Sciences, Peace and Conflict Studies (IJMSSPCS), 4(3) :241 – 253.

[7] Gitari, Z., Jun Long and Hanyurwimfura, D. 2015. A Lexicon-based Approach for Hate Speech Detection. April 2015 International Journal of Multimedia and Ubiquitous Engineering 10(4):215-230 DOI:10.14257/ijmue.2015.10.4.21

[8] Guyal, K. 2020. Top 4 Types of Sentiment Analysis & Where to Use. Retrieved from https://www.upgrad.com/blog/types-of-sentiment-analysis/ on 26-01-2023

[9] Haralambous, Y. and Lenca, P. 2014. Text Classification Using Association Rules, Dependency Pruning and Hyperonymization. Published in DMNLP@PKDD/ECML 28 July 2014.

[10] Ugwumba, A. 2008. Prevalence of Herdsmen and Farmers Conflict in Nigeria. International Journal of Innovative Studies in Sociology and Humanities, 3(1).

[11] Karause, G. and Grasseggles, A. 2010. Assessment of the Effect of Farmers-Herdsmen Conflicts on National Integration in Nigeria. International Journal of Humanities and Social Sciences, 8 (10), 118—128. DOI: 10.30845/ijhss.v8n10p13.

[12] Mehdad, Y. and Tetreault J. 2016. Do Characters Abuse More Than Words? Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue. DOI:10.18653/v1/W16-3638.

[13] Statista. 2019. Considerations on the use of instagram as a pedagogical tool under the view of contextualized education. Journal **o**f Creative Education**,** Vol.12 No.7, July 27, 2019.

[14] Alao, D. O., Shaibume, B., Ogunwemimo, T., Alao, E. M. and Ogunwemimo, O. 2019. Herdsmen/ Native Farmers' Violence in Benue State and Food Security in Nigeria. Mediterranean Journal of Social Science, 10(6).

[15] Kah A.E and Zeroual, I. 2022. "Sentiment analysis of students' Facebook comments toward university announcements," 2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS), Bandung, Indonesia, 2022, pp. 1-5, doi: 10.1109/NISS55057.2022.10084994.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 20 - 24, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1004

Development of a Museum Virtual Tour Based on the Art Steps Application of the Lae Meang Old Tomb Site, Pakpak Bharat, Indonesia

Ika Purnamasari
Departement of History Education
Universitas Negeri Medan
Medan, Indonesia

Ichwan Azhari
Departement of History Education
Universitas Negeri Medan
Medan, Indonesia

Pidia Amelia
Departement of History Education
Universitas Negeri Medan
Medan, Indonesia

Abstract: The museum model in the modern era has experienced very rapid development. Initially, museums could not be separated from buildings and archaeological objects as collections. However, now the museum has been presented in a new form, namely the development of a virtual museum tour using various applications. In this research, the application used is the ArtSteps app. The objects or collections that will be displayed in this virtual museum are various types of archaeological remains found at the Old Lae Meang Grave Site, Mahala Village, Pakpak Bharat Regency, North Sumatra Province, Indonesia. At this site, 45 old graves were found which could be evidence of the entry and development of Islam in Pakpak Bharat Regency. This site also has high importance for history, science, education, religion, and culture. The benefit of this research is to understand historical heritage in Indonesia which is difficult to reach, so with the help of a virtual museum tour, you can save time, costs and present historical heritage with more realistic images.

Keywords: Virtual Tour, Museum, ArtSteps, Archaeology, Tombstone

1. INTRODUCTION

It cannot be denied that the impact of the Covid-19 pandemic is changing the way society interacts. At first, everything was meeting face to face, but suddenly it changed to using online media. The Covid-19 pandemic has also had a changing impact on the world of museums. During Covid, people's movement was limited so they couldn't visit the museum directly. Activities at the museum were suddenly paralyzed, due to the loss of visitors. To answer the challenges caused by the COVID-19 pandemic, museums have developed various strategies so that the museum world remains alive and can be enjoyed by the public. One of them is changing the exhibition directly in the museum room (offline) into an online exhibition [3]. This can be implemented using the Artsteps application.

Artsteps technology makes it easy for users to create three-dimensional virtual exhibition spaces for interior and exterior spaces and can use virtual reality (VR) to give a more real impression. Creating a virtual museum using a VR web such as the artsteps.com application is very suitable for designing virtual museum tours [1]. Pilliang (2004) states "The growth of information technology, especially the internet, has increased the need for real-time style or lifestyle, namely a pattern that requires everything to be done through virtual channels, which is done in aesthetics". In this way, information technology, especially the internet, has become mandatory in social life because it has become a lifestyle.

Currently, there are various kinds of websites on the internet, one of which is a gallery website. Technologically, according to Amri Yahya, a gallery is a place for displaying art objects or other cultural objects (including historical items) that are strictly selected and selected by a team or an expert who truly has quality. This is needed as a guarantee of quality [6]. Reflecting on this explanation, we can see the meaning of a gallery which is also a form of the character of a museum. Museum collections that have been displayed conventionally in exhibition vitrines can be transformed into virtual exhibitions. Even archaeological objects on the site can also be collected in virtual museum exhibitions.

The Lae Meang Ancient Tomb Site is located very far inland, with this virtual museum service it is possible to exhibit it in virtual form too. Ancient tombstones with various distinctive and unique typologies are part of the collection in this virtual museum. The archaeological remains found at this site are tombstones with various typologies and ornamental varieties. This study is a preliminary review in identifying the existence of Hindu influences on tombstones at the Lae Meang Site. Through the stages of archaeological research by analyzing artifactual remains, this research revealed that elements of Hindu tradition are found in several typologies of phallus-shaped tombstones (phallus) and flattened at the Lae Meang Site. Although it dates back to the early 20th century, the typology of tombstones such as those at the Lae Meang Site has never been found before at other Islamic sites in Indonesia [4].

The archaeological remains of the Lae Meang Ancient Tomb Site can now be presented in a special room by following developments in information technology. Currently, there are many virtual galleries or gallery websites that present 3-dimensional exhibitions online or virtually, Artsteps.com is one of the gallery web pages that can present works of art virtually with high resolution, allowing visitors to feel like they are at a real exhibition even though they are Visitors only watch art exhibitions virtually via smartphone or laptop. The functions of the ArtSteps application include; (1) a Web gallery that presents virtual works of art in high resolution, (2) Helping creators create exhibitions, events, and promotions with a 3-dimensional realistic room concept, (3) Art room features for artists, (4) Helping in Open a web page to do work in the field of artwork.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 20 - 24, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1004

2. METHOD

This research uses a research and development approach. The research and development of the ADDIE model has procedures structured with systematic sequences of activities at each development stage. In this research, there are five stages, namely (1) Analysis, (2) Design, (3) Development, (4) Implementation and ending with stage (5) Evaluation [5].

This research uses an archaeological approach to identify and map Islamic archaeological remains in Pakpak Bharat Regency, especially in Lae Meang Village. In this approach, the method used consists of four stages: 1) field observation or survey; 2) data identification; 3) data analysis; and 4) synthesis. When carrying out a survey, some archaeological data must be collected, including artifacts, ecofacts, features, and traces of ancient building structures which are thought to be evidence of traces of past human activity. The data expected to be obtained from this surface survey is in the form of a general description of the geomorphology of the area as well as various types of artifacts or ecofacts. Next, the data identification process is carried out to determine the morphological form of archaeological remains in the form of size, shape or typology, material, color, decorative patterns, and other data so that it can be analyzed to determine the relative period of use and use in the past.

3. RESEARCH RESULT

Procedures for Development of a Virtual Museum Tour

The development model used in compiling a virtual tour of the Lae Meang Old Tomb Site museum is the ADDIE development model. This model has five research stages, namely, Analysis, Design, Development, Implementation, and Evaluation [5].

3.1. Analysis

The analysis carried out at this stage was to analyze the needs of the community represented by 60 students from the Department of History Education, UNIMED, to measure the level of interest in developing a virtual museum tour model. Based on a questionnaire distributed to students taking Archeology courses, 60 respondents found that they wanted the development of a virtual museum to at least have innovative elements, provide a real experience, contain 3-dimensional elements, and be equipped with illustrations, 3D animation, and video. The following is a recapitulation of the results of the needs analysis regarding this matter.

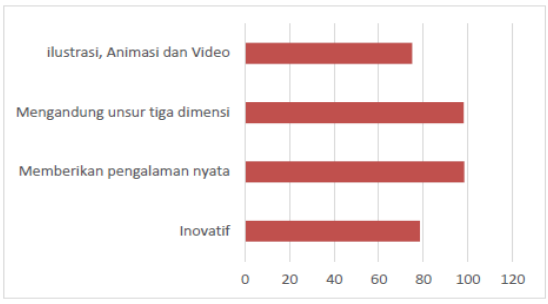


Figure 1. Percentage of Student Needs for Virtual Museum Tour Services

In this stage, data identification is carried out on objects that will become material in the virtual museum. The data is in the form of tombstones at the Lae Meang Old Tomb Complex Site, Pakpak Bharat. In an archaeological study carried out in previous research in Lae Meang Hamlet, Mahala Village, an Islamic archaeological site was discovered, namely an ancient Islamic tomb complex. From this study, it was identified that there were at least 45 graves or around 90 tombstones. Interestingly, the tombstone typology at the Lae Meang Site is different from the tombstone typology at other Islamic sites [4]. Lae Meang is one of the names of the hamlets in Mahala Village, Tinada District, Pakpak Bharat Regency. Geographically, Lae Meang Hamlet is located at coordinates 2 o 15' 32" N and 98 o 31' 12" E. Lae Meang Hamlet has quite a distinctive geographical character because it is located on a hilly plateau with a height of 600 meters above sea level and is dominated by traditional forests from the traditional land of the Solin clan. For this reason, the customary forest is still very sustainable, and only 8% has begun to be used as community plantation land.

Figure 2. Lae Meang Old Tomb Complex Site

The geographical characteristics of being in the hills or highlands make it different from the tomb complexes in coastal areas or lowlands. Apart from that, the presence of the Pakpak ethnic group as an ethnic supporter of the tomb complex also contributes to giving color to the decorative patterns on the tombstones. Although from the analysis that has been carried out, it is known that the age of the Lae Meang Site is thought to be not too old, namely dating from the early 20th century AD.

In local tradition, it is stated that in the past the Lae Meang area was thought to be the location of traditional or lebbuh settlements by the Pakpak Bharat community, especially those from the Solin clan group. Lebbuh itself is led by a Pertaki , namely the eldest son of a clan. Therefore, in the Lae Meang area, two other tombs were also found at the foot of the hill which local people believe are the tombs of the Pertaki Lebbuh Solin clan. Apart from that, several umpak stones, or in the local language called cepu were also found in the Lae Meang area which functioned as foundation stones for a building. The cepu stone is thought to be a stone from the Sapo Jojong building of the Solin clan in that area in the past. So in terms of customary legitimacy in Pakpak Bharat, the Lae Meang area is considered a forest or traditional area of the Solin clan. In general, the current condition of the Lae Meang Site is in the middle of the Solin clan's traditional forest and is about 1 km from the nearest settlement [4].

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 20 - 24, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1004

3.2. Design

The next stage is entering the design stage which consists of several activities including developing initial ideas, carrying out data inventory in the field, namely the need to take photos of tombstones as a collection for the virtual museum to be designed, selecting content in the museum collection along with other historical data. The design stages in this research are as follows:

3.2.1. Create Flowcharts

The next stage is creating a program design in the form of a flowchart. This flowchart contains the development of concepts and data that will become content or collections in the virtual tour of the museum. Flowcharts consist of opening, content, and closing aspects.

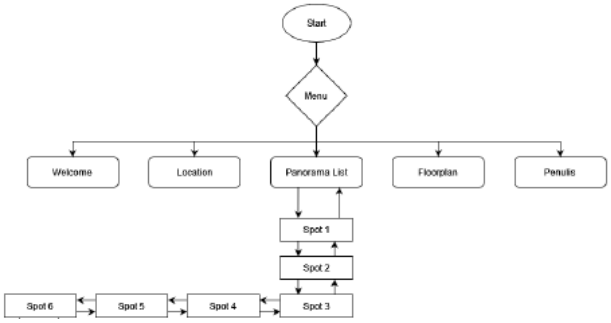


Figure 3. Virtual Museum Flowchart for the Lae Meang Old Tomb Complex Site

3.2.2. Create Storyboards

Storyboards are used to create a web display design for the photography gallery being developed. The storyboard contains the stages and materials that will be displayed on the virtual tour of the museum. The material that will become the museum collection comes from archaeological research that was carried out at the Lae Meang Tomb Complex Site in 2022. In the archaeological study that was carried out in previous research in Lae Meang Hamlet, Mahala Village, an Islamic archaeological site was discovered, namely a tomb complex. ancient Islam. From this study, it was identified that there were at least 45 graves or around 90 tombstones. Interestingly, the tombstone typology at the Lae Meang Site is different from the tombstone typology at other Islamic sites[4]. Based on the results of this research, the material in the virtual tour museum storyboard is as follows:

No	Tampilan Koleksi Museum Virtual	Narasi
1		***Tipe Batu Nisan Aceh*** Pada kompleks makam Lae Meang ditemukan 2 batu nisan dengan tipologi batu nisan aceh. Nisan ini memiliki hiasan yang sangat kaya, antara lain berupa hiasan geometris, floris (berbentuk bunga-bungaan) dan motif khas Aceh. Tipologi batu nisan spiral diindikasikan berjenis kelamin laki-laki. Diduga berasal dari abad ke-19 M
2		***Tipe Silindrik Lingga;*** Untuk wilayah Sumatera Utara, batu nisan dengan tipologi lingga baru hanya di temukan di Situs Lae Meang, Pakpak Bharat. Tipologi batu nisan spiral diindikasikan berjenis kelamin laki-laki. Terdapat inskripsi: "Ari *Arba* Bulan Maret Tahun 1928 nun …… Allah"
3		***Tipe Pipih Antromorpik;*** Batu nisan dengan tipe pipih berundak seperti ini juga baru hanya di temukan di Situs Lae Meang, Pakpak Bharat. Menariknya motif hias pada nisan ini hampir seluruhnya dikenal dalam tradisi Papak, sepeti gerga perkupkup manun, gerga perbunga rintua, gerga perbunga paku dan gerga perbunga kimbang. Tipologi batu nisan pipih diindikasikan berjenis kelamin perempuan. Terdapat inskripsi: "Simuna Marga Tumagir, Hari 18 Bulan Ramadan Tahun 1928, Kira Kira Umurnya 20 Tahun, Ini Perempuan Yang Mulia"
4.		***Tipe Pipih Akolade;*** Batu nisan dengan tipe pipih akolade dengan motif hias seperti baru ditemukan di Situs Lae Meang, Pakpak Bharat. Tipologi batu nisan pipih diindikasikan

		berjenis kelamin perempuan. Estimasi berasal dari abad ke-20 M

3.3. Development

The development stage is the stage of producing, purchasing or revising the materials needed to achieve development goals. In the development stage, there are several activities including the process of developing content (text, images, 3D objects, audio, and video) which is designed in the form of flowcharts and storyboards, collecting data from various sources, in the form of photos, videos and documentation regarding the Tua Lae Tomb Complex Site. Yes. Photos and images of the virtual museum collection were taken directly by the research team when they visited the Lae Meang Site on 15 – 18 May 2022. These photos were then developed into virtual museum products using the ArtSteps application.

In the process of developing a virtual museum tour using the Artsteps.com application, the steps that must be taken include:

3.3.1. Create an ArtSteps.com account and log in.

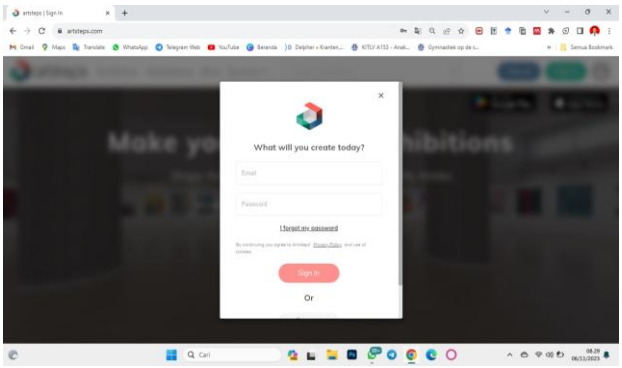


Figure 4. Stage 1

3.3.2. Create a room layout according to the theme

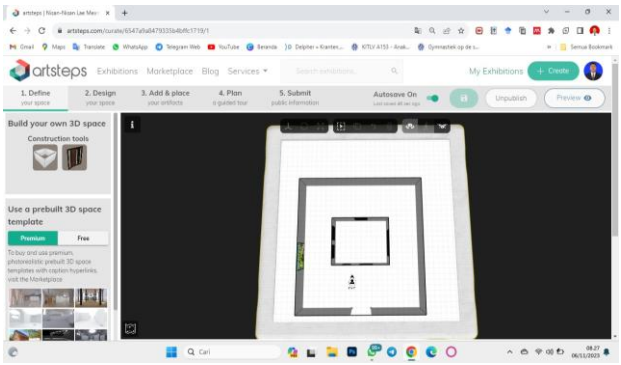


Figure 5. Stage 2

3.3.3. Upload photos and arrange photo layouts according to a predetermined theme

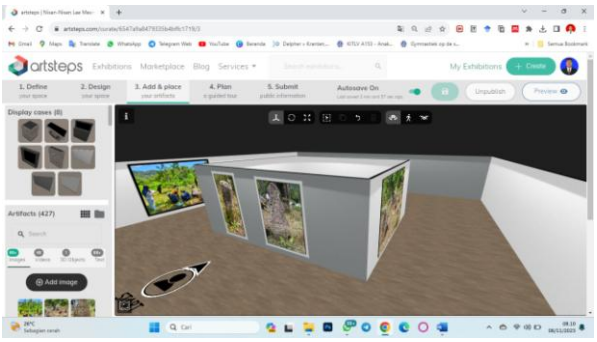
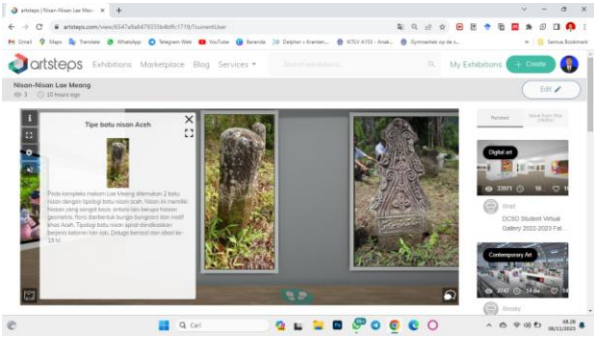


Figure 6. Stage 3

3.3.4. Provides supporting narrative to virtual museum collections



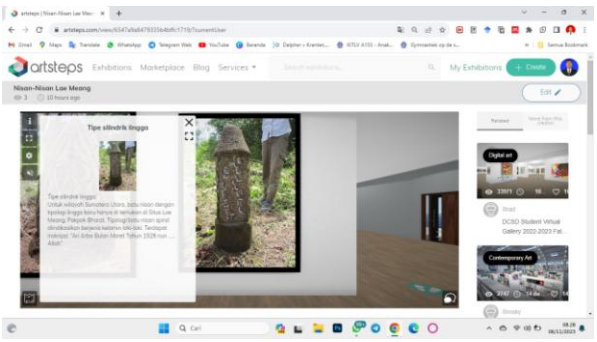


Figure 7. Stage 4

3.4. Implementation

After the analysis, design, and development stages have been completed, the next stage is the implementation stage. At this stage, a Virtual Tour of the museum was implemented using ArtSteps.com for students of the Department of History Education, UNIMED, via the Zoom application.

3.5. Evaluation

After carrying out the analysis, design, development, and implementation stages, the next stage is the evaluation stage. This stage consists of several steps, including language evaluation, material (content) evaluation, media evaluation, and field trials.

No	Aspek	Validator (*Expert*)	Skor	Kategori
1	Media	KW	91.81 %	Sangat Valid
2	Materi	HD	70.90 %	Valid
3	Bahasa	FAR	89.69 %	Sangat Valid
Rerata			**84.14 %**	**Sangat Valid**

(Sumber : Hasil Pengolahan Data Primer 2022)

. Figure 8. Evaluation Results of the Virtual Tour Museum of the Lae Meang Old Tomb Complex Site

4. CONCLUSION

The results of this research conclude that it is very important to develop virtual museum tours, especially in the field of in-situ archeology such as tomb complexes. The presence of a virtual museum tour of these archaeological sites helps people to visit an archaeological site without having to come directly to a location that is usually in a remote, remote, remote area and has difficult access, such as the Lae Meang Ancient Tomb Site in Pakpak Bharat. The Lae Meang Ancient Tomb Site can now be presented directly to people wherever they are simply by accessing the virtual tour of the museum which has been prepared via the ArtSteps application.

5. ACKNOWLEDGMENTS

We would like to thank LPPM Universitas Negeri Medan for funding our research and all participans and supervisors that contributed to the work in this study.

6. REFERENCES

[1] Alfarizi, M., & Yugopuspito, P. (2020). Pengembangan Museum Virtual Reality Berbasis Inkuiri dalam Pembelajaran Pendidikan Pancasila dan Kewarganegaraan. Jurnal Pendidikan, 21(2), 94-103. https://doi.org/10.33830/jp.v21i2.974.2020.

[2] Effendi, Empy dan Hartono Zhuang, (2005). “E-learning Konsep dan Aplikasi”. Yogyakarta: Penerbit Andi

[3] Hasyim, N., & Senoprabowo, A. (2019). Perancangan Ruang Pamer Digital dalam Media Virtual Reality sebagai Upaya Menyediakan Ruang Pamer Interaktif. Gestalt, 1(1), 103-112. https://doi.org/10.33005/gestalt.v1i1.24.

[4] Purnamasari, Ika, dkk. (2022). “Identifikasi Tinggalan Arkeologi Islam dan Strategi Pengelolaan Wisata Situs Lae Meang, Kabupaten Pakpak Bharat”. Yogyakarta: Pena Persada

[5] R. M. Branch, Instructional Design: The ADDIE Approach. Boston, MA: Springer US, 2009

[6] Siswanto, Aris Budi. 1996. Galeri Seni Lukis di Yogyakarta, Landasan Konseptual Perancangan, Laporan Tugas Akhir Jurusan Teknik Arsitektur Fakultas Teknik Sipil dan Perencanaan Universitas Islam Indonesia Yogyakarta

[7] Sugiyono, P. D. (2017). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.

International Journal of Computer Applications Technology and Research
Volume 12–Issue 11, 25 - 28, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1211.1005

Preparation of a Thematic Map of Agrarian Conflict Based On Digital Database of Conflict Distribution in North Sumatra

Tappil Rambe
Departement of History Education
Universitas Negeri Medan
Medan, Indonesia

Denny Setiawan
Departement of PPKN Education Universitas Negeri Medan
Medan, Indonesia

Pidia Amelia
Departement of History Education
Universitas Negeri Medan
Medan, Indonesia

Abstract: Agrarian conflict in North Sumatra Province has a very long history. Starting from the mid-19th century to the current modern era, it takes place in eight areas such as Batubara Regency, Asahan Regency, Deli Serdang Regency, North Padang Lawas Regency, Serdang Bedagai Regency, Simalungun Regency, South Tapanuli Regency, and Toba Samosir Regency. Therefore, it is important to develop a database of agrarian conflicts in North Sumatra Province in the form of a thematic map of conflict distribution so that the history and existing conditions of existing conflicts can be easily understood. The development of this digital database also aims to ensure that existing conflicts can be identified systematically, with the aim that conflict resolution policies can be made optimally. Eight districts in North Sumatra which have a history of quite complex agrarian conflicts, some of which have received solutions that are acceptable to all parties. However, in several other districts, agrarian conflicts continue to this day. The large number of actors who have their own interests in a land and the absence of a common opinion results in continuous conflict. Therefore, it is very important for all data on agrarian conflicts that have occurred to be compiled in a digital database, namely in the form of a thematic map, so that the formula for conflict resolution over land in an area can be prepared more comprehensively.

Keywords: Digital Database, Agrarian Conflict, Thematic Map

1. INTRODUCTION

The agrarian conflict in North Sumatra Province has a very long historical background. History records that agrarian conflicts in North Sumatra began to emerge in the mid-19th century when foreign plantation companies began to establish themselves in the East Sumatra region (currently part of North Sumatra Province). The beginning of plantation companies in North Sumatra began in 1864, when Jacobus Nienhuys, a Dutch foreign private entrepreneur, tried to plant tobacco on Deli land after obtaining permission from the Sultan of Deli. Nienhuys' business was successful where Nienhuys harvested 50 bales of tobacco. This marks the presence of Deli tobacco which is growing rapidly in East Sumatra, which in a short time has become very popular in the European tobacco market as the best cigar wrapper in the world. The widespread opening of tobacco plantations immediately transformed empty lands or jungles in East Sumatra into areas with the most productive economy and provided large profits to plantation or plantation entrepreneurs (Stoler, 2005, p.2).

Several recorded plantation companies include Deli Batavia Maatschappij (1875), Arensburg (1876), Amsterdam Deli Compagnie (1979), Rotterdam Deli Maatschappij (1881) and Harrison & Crossfield (1906). It was recorded that from 1863 – 1906 there were around 170 plantations. One factor in the rise of tobacco plantations is, among other things, the granting of permits or concessions by the Sultanate of Deli to use land or land in its "territory". The Sultanate of Deli was lured by even more fantastic economic profits by attracting it to the capitalist economic system, among other things, by controlling land (Anderson, 1862).

The impact of unilateral control of community land into plantations led to the emergence of agrarian conflicts which manifested in several battles or wars such as the Single War in 1872 - 1879 (Pelzer, 1985, p. 62-63). In 1873, there was also a battle between the Karo Batak or Datuk Penghulu Batak Karo and the Dutch because the Deli Sultanate unanimously determined land use, economic and political spaces in Karo (Ratna, 2004).

The political phenomenon that occurred in 1872 showed that there was enormous turmoil in rejecting intimidation or colonialism in both the political and economic fields. The presence of plantation companies actually destroys the existence and authority of local residents in managing their economy which originates from pepper plantations. At the same time destroying land which was originally a pepper plantation area which was then converted into a tobacco plantation.

During the Japanese occupation, empty land was planted with food crops to support the war waged by Japan, which was implemented by inviting immigrants to plant empty plantation land, resulting in the existing order being destroyed. Many tract lands are occupied by illegal immigrants. After the 1946 Social Revolution, several plantation areas were controlled by people's troops. The control of plantations by these irregulars then turned into an arena of attraction in fighting over sources of economic production (Agustono, 1997). This illegal cultivation of tobacco plantation lands is becoming increasingly unstoppable by the government. Meanwhile, the local people (indigenous

residents) themselves do not involve themselves in reducing the land for tobacco plantations because they still adhere to customary law and only manage tract lands.

This long historical background shows evidence that North Sumatra has very strong agrarian conflict roots and continues to develop in the contemporary context. It is recorded that 8 districts in North Sumatra Province have strong agrarian conflict data with different characters. Some of them have found solutions, but some of them still continue to struggle in the midst of conflict. Therefore, it is important to develop existing conflict data into conflict distribution maps that can be accessed digitally. From the existence of this digital database in the form of a map, it is hoped that all interested stakeholders can formulate an appropriate conflict resolution.

2. METHOD

By using quantitative descriptive methods, data analysis is carried out through interpretation based on intellectual understanding built by empirical experience, interpretation and data analysis are carried out in the following steps:

a) Data collection, through documentation techniques to obtain secondary data as well as interview and observation data to obtain primary data.
b) Data assessment by paying attention to the principles of validity, objectivity and reliability. For this reason, the following procedure is carried out:
   - Categorize primary data and secondary data with relevant recording
   - Evaluate the data that has been obtained with the aim of controlling whether the data is relevant for use.
c) Interpretation and presentation of data is carried out by analyzing data and facts through careful intellectual understanding and must be equipped with a set of relevant theories.
d) Inference, namely drawing conclusions based on interpretation and analysis of data in the form of compiling a digital database of agrarian conflicts in North Sumatra.

3. RESEARCH RESULT

Land Disputes Are the Root of Agrarian Conflict

North Sumatra Province has an area of 71,680.68 km2. In its development, recently it has experienced many changes, especially in terms of land use. This occurs in line with the increasing implementation of development and the economy in the region, causing an increase in land conversion changes which result in land use no longer being appropriate as determined in accordance with the principles of sustainable development.

Land disputes that often arise in people's lives are partly due to struggles over land rights which result in the destruction of harmonious social relations. In customary law communities, disputes often occur regarding customary lands, including ulayat land. The causes of ulayat land disputes include: lack of clear boundaries regarding customary land, lack of awareness of customary law communities, and the role of traditional heads in customary law communities.

The agrarian conflict in North Sumatra is one of the largest in Indonesia, KPA (Consortium for Agrarian Reform) noted throughout 2020. Agrarian conflict is interpreted as conflict related to land. There are several factors that cause this conflict to occur, including issues of land control or land disputes and competition for natural resources. Agrarian conflicts arise as a result of gaps in agrarian resources, for example plantation land or customary land. In agrarian conflict, many parties are involved in it and it is characterized by the presence or implementation of regulations and decisions that are not balanced so that the conflict becomes more complex (Sumarjono, 2013).

North Sumatra Agrarian Conflict Digital Database

Dietz said that in general the symptoms of conflict in natural resource relations are rooted in conflicting claims regarding three things, namely: (1) who has the right to control agrarian resources and the natural wealth contained therein; (2) who has the right to utilize agrarian resources and natural wealth; (3) who has the right to make decisions regarding the control and utilization of agrarian resources and natural wealth (Dietz, 1988. p. 8).

The distribution of conflicts and the causes of agrarian conflicts in 8 districts in North Sumatra can be seen in the recapitulation data on land conflicts in Sumatra which have been compiled in a digital database.

No	Lokasi	Luas Lahan	Aktor	Faktor Konflik
1	Kabupaten Asahan	1650 Ha	Masyarakat PTPN II Kelompok Tani Mekar Jaya	• Pemerintah tidak tegas • Pemerintah belum mengambil keputusan penerapan hukum
2	Kabupaten Batubara	5720 ha	Masyarakat Kelompok tani Perusahaan	• BPN tidak menyelesaikan masalah sertifikasi tanah • Pemeirnatah tidak memiliki kesamaan pandangan • Pengambilan dan perebutan

				lahan
3	Kabupaten Deli Serdang	12.570 ha	Masyarakat Kelompok Tani PTPN II Perusahaan Forum Komunitas Indonesia Bersatu	• Pengambilan dan perebutan lahan • BPN belum memastikan batas wilayah yang tepat • Adanya pihak-pihak yang ingin memperkeruh suasana, seperti mafia tanah • Adanya aturan hukum yang tidak dapat diterapkan di lapangan
4	Kabupaten Padang Lawas Utara	9020 h	Masyarakat Perusahaan	• Belum ada kejelasan dari pemerintah • Pihak BPN masih mengadakan pengukuran • BPN tidak transparan dalam pengurusan surat
5	Kabupaten Serdang Bedagai	6780 ha	Majelis Adat Melayu Masyarakat PTPN II	• Pemerintah belum mengambil keputusan penerapan hukum • Adanya peraturan yang belum dapat diterapkan di lapangan • BPN tidak menyelesaikan masalah sertifikasi tanah •
6	Kabupaten Simalungun	1570 ha	Masyarakat Perusahaan Komunitas	• Pengambilan dan perebutan lahan • BPN belum memastikan batas wilayah yang tepat • Adanya pihak-pihak yang ingin memperkeruh suasana, seperti mafia tanah
7	Kabupaten Tapanuli Tengah	2030 ha	Masyarakat Kelompok Tani Perusahaan	• Pemerinatah diduga membackup beberapa perusahaan yang HGUnya belum diperpanjang • Administrasi pertanahan kacau
8	Kabupaten Toba	2367 ha	Masyarakat Perusahaan Kertas	• Pemerinatah diduga membackup beberapa perusahaan yang HGUnya belum diperpanjang • Administrasi pertanahan kacau

Table 1. Database of the Extent and Distribution of Agrarian Conflict in North Sumatra (Source from Interviews and Analysis of Various Sources, 2022)

The database on the extent and distribution of conflict shows that the district with the greatest conflict conditions is Deli Serdang Regency, namely with an area of 12,570 ha spread over 35 cases in various sub-districts. Meanwhile, the district with the smallest conflict conditions is Simalungun Regency with an area of 1,570 ha spread over only 6 cases.

Based on the table above, it can be seen that the oldest land conflict problem, which started from 1942 to 2022, is the land conflict problem between the V Jati Mulyo Hamlet Community Group, Tebing Tanjung Selamat Village, Kec. Padang Tualang with a conflict area of 23.6 ha, this problem has not yet been resolved, the reason is that the community is still ensuring the completeness of documents from residents.

Based on the table above, it can also be seen that the parties involved in land conflicts can be grouped into several categories, namely:

1. Land conflicts between community groups and state-owned companies and private companies.

2. Land conflicts between farmer groups and state-owned companies and private companies.

3. Land conflicts between traditional councils and state-owned companies.

Based on this fact, society is the party most disadvantaged, especially farming communities who, with all their limited abilities, are always burdened by economic poverty, lack of legal ability to fight for their rights, lack of access to support and a myriad of other weaknesses, putting them in a very difficult position and In fact, it is not uncommon for them to become victims of tyranny and greed by the management of plantation companies (PTPN & Private Plantations) who, with their abilities, easily use the hands of those in power (Government Bureaucrats, Security Authorities/Polri, TNI, Judicial Institutions, even thugs) to beat and muzzle the community. poor poor farmers who fight for their rights. The land conflicts tabulated in the table are land conflicts that have not been resolved. Therefore, to find out how to resolve the land conflict, it is done by comparing it with other cases, especially those related to land conflict issues.

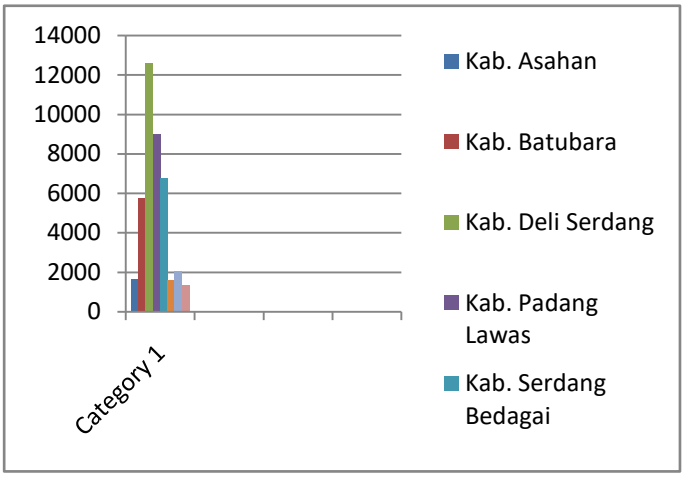


Diagram 1. Calculation of the Extent of Agrarian Conflict in North Sumatra Based on Digital Agrarian Conflict Data.

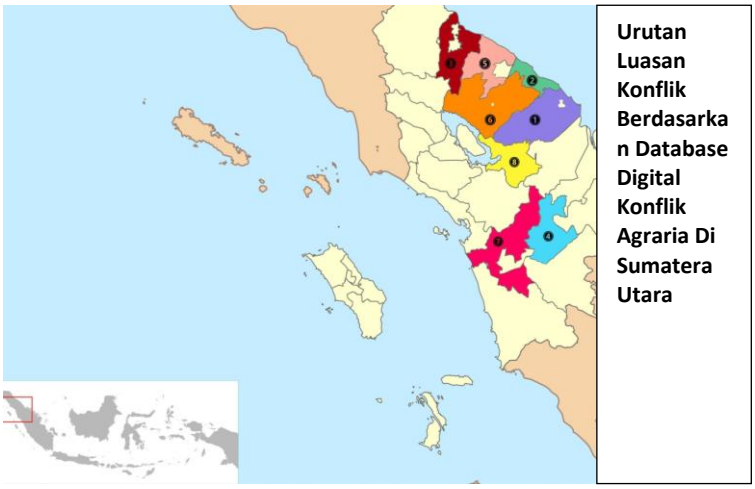


Figure 1. Distribution of Agrarian Conflict in North Sumatra Based on Digital Database

Conflicts of interest arise between the parties involved, where each of them acts as a sharecropper (horizontal conflict) and there is an interest on the part of the plantation entrepreneur to defend their concession area or Erfacht/HGU rights, which has economic nuances to gain profit (vertical conflict). Apart from that, there is an interest of entrepreneurs to regulate the cultivators on plantation land. In order to develop the economy as a newly independent country, the government tends to protect the plantation entrepreneurs rather than the farming people. This fact ultimately gave rise to disputes between the cultivators, the government and plantation entrepreneurs.

4. CONCLUSION

Based on the results of statistical data analysis, the factors that cause conflict are used as research objects, namely community, government, law, culture, law enforcement, land administration bodies simultaneously influence the number of land conflicts in North Sumatra. This means that these six factors can simultaneously influence the size of the number of land conflicts in North Sumatra.

5. ACKNOWLEDGEMENTS

We would like to thank LPPM Universitas Negeri Medan for funding our research and all participans and supervisors that contributed to the work in this study.

6. REFFERENCES

1) Anderson, John. 1971. *Mission to the Eastcoust of Sumatra in 1823*. Kuala Lumpur: Oxford University Press.
2) Armani, Hamidah, dkk. 2020. *Konflik Agraria Pada Masa Revolusi Sosial Di Sumatera Timur 1946-1955*. Jakarta: Jurnal Historia.
3) Creswell, J.W. 1998. *Qualitative Inquiry and Research Design: Choosing among Five Tradition*. London: Sage Publications.
4) Nurhasim, Moch. 2008. *Konflik dan Integrasi Politik Gerakan Aceh Merdeka*. Yogyakarta: Pustaka Pelajar.
5) Pelzer, Karl J. *Toean Keboen dan Petani, Politik Kolonial dan Perjuangan Agraria di Sumatra Timur, 1863-1947*. Jakarta: Sinar Harapan, 1985.
6) Ratna. Aceh, "*Deli dan Perang Sunggal*". *Papers* in Perang Sunggal Seminars, Medan, 2004.
7) Stoler, Ann Laura. 2005. *Kapitalisme dan Konfrontasi di Sabuk Perkebunan Sumatera, 1870 – 1979*. Jakarta: Karsa.
8) Sumarjono. 2013. *Konflik Agraria: Pengelolaan Sumber Daya Air*. Yogyakarta: APMD Press