Development and Implementation of an Automated Ticketing System for Uganda Electricity Distribution Company Limited (UEDCL) Customers.

PROJECT REPORT

Submitted November 2025

BY MUTEBI SAMUEL

REGNO: 2025/HD/05/26356U

STUDENT NO: 2500726356

COURSE: Master of Science in Computer Science

COURSE Unit: MCS 7103: MACHINE LEARNING

LECTURER: DR. Lillain Muyama

SAMUEL MUTEBI 2025/HD/05/26356U 2500726356

Table of Contents

List of Abbreviations

UEDCL: Uganda Electricity Distribution Company Limited

SVM: Support Vector Machine

TF-IDF: Term Frequency–Inverse Document Frequency

SMOTE: Synthetic Minority Over-sampling Technique

NLP: Natural Language Processing

Project Background:

The developed ticketing system automates the assignment of field technicians to incoming customer complaints in utility services (Case study UEDCL). Traditional systems rely on manual or static database mappings for technician assignment, which are inefficient and error-prone as complaints volume grows. This system harnesses machine learning, specifically a Support Vector Machine classifier, to predict the most suitable technician based on historical ticket data, thereby improving accuracy and operational efficiency. The system loads complaint data, prepares and trains a Support Vector Machine (SVM) model on complaint texts to predict the technician who will resolve the reported issue, and integrates this with a database ticketing system.

System Functionality:

- The system loads a dataset from an Excel file (Data extracted from UEDCL complaints Management system) and cleans/normalizes some fields.
- It encodes the technician labels and vectorizes complaint texts using Term Frequency–Inverse Document Frequency (TF-IDF).
- The system splits data into training and test sets.
- It trains an SVM classifier pipeline to predict the technician responsible for resolving a complaint.
- After training, the system evaluates model performance using confusion matrix, overall accuracy, and recall per class.
- The system sets up an SQLite database for managing tickets with fields for user information, complaint details, technician assignment, etc.
- Functions **generate unique ticket** references, predict technicians using the model, and add new tickets to the database with validation.

- For interactive fields, the system uses Jupyter interactive widgets to allow users to submit tickets, update statuses, and perform ticket number searches.
- Quality is evaluated with a confusion matrix, accuracy, and balanced recall scores

Dataset Used:

The dataset used is provided as "DATASET-4.xlsx," containing historical records of tickets with fields such as REGION, DISTRICT, CUSTOMER NAME, COMPLAINT_TYPE (Categorical field specifying the type of complaint), COMPLAINT (text description of the issue), RESOLVED_BY (technician), STATUS, and REFERENCE_NO. The dataset comprises thousands of tickets from multiple regions and districts, with diverse complaint types and technician assignments. This data forms the training foundation that enables the system to learn patterns and assign technicians automatically for new tickets

REGION	DISTRICT	CONTRACT_NO	CUSTOMER_NAME	COMPLAINT_TYPE	COMPLAINT_GROUP	COMPLAINT	RESOLVED_BY	STATUS	REFERENCE_NO
CENTRAL	BOMBO	102051690263	ALLEN TAMALENALONGO	Meter Replacement	EQUIPMENT_COMPLAINT	FAULTY METER REPLACEMENT	SEPIRIYA SEKABEMBE	CLOSED	R2004241200533
NORTH EAST	GULU	100030861427	AKELLO PEACE	Meter Replacement	EQUIPMENT_COMPLAINT	faulty new meter	NELSON KENNEDY	CLOSED	R2004250100173
CENTRAL	BOMBO	102064549237	KAMULEGEYA SENTONGODEMBE	Meter Replacement	EQUIPMENT_COMPLAINT	FAULTY METER RPELACEMENT	SEPIRIYA SEKABEMBE	CLOSED	R2004250100307
WESTERN	KAGADI	100027625142	MATAYO KATEMBA	Meter Replacement	EQUIPMENT_COMPLAINT	faulty meter	SAUDA NAMATOVU	CLOSED	R2004250100317
NORTH EAST	MBALE	102069665069	MARTIN NETONGE	Meter Replacement	EQUIPMENT_COMPLAINT	FAULTY METER	ELIZABETH NAMUGANGA	CLOSED	R2004250100449
NORTH EAST	MBALE	100002673030	MUTAMBO FRED	Meter Replacement	EQUIPMENT_COMPLAINT	FAULTY METER	ELIZABETH NAMUGANGA	CLOSED	R1000241200013
NORTH EAST	MBALE	100020565675	MUTAMBO FRED	Meter Replacement	EQUIPMENT_COMPLAINT	FAULTY METER	ELIZABETH NAMUGANGA	CLOSED	RCALL241200460
NORTH EAST	JINJA	102006770564	ST.NOA MAWAGALI SEC SCH	Rebilling	BILLING_REQUEST	To correct the September readings	IRENE ESAETE	CLOSED	RCALL250200154
NORTH EAST	GULU	102068658693	ODIYA FINNY	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	No power but the rest have in kabalega road	NELSON KENNEDY	CLOSED	R1000241200007
NORTH EAST	GULU	100002403150	ODIYA FINNY	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	No power but the rest have in kabalega road	NELSON KENNEDY	CLOSED	RCALL241200014
NORTH EAST	GULU	100020941300	ODIYA FINNY	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	No power but the rest have in kabalega road	NELSON KENNEDY	CLOSED	RCALL241200018
NORTH EAST	GULU	100011145354	ODIYA FINNY	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	No power but the rest have in kabalega road	NELSON KENNEDY	CLOSED	RCALL241200021
NORTH EAST	GULU	100002115515	ODIYA FINNY	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	No power but the rest have in kabalega road	NELSON KENNEDY	CLOSED	RCALL241200181
NORTH EAST	GULU	102064104557	BENSON OYATOLIT	Meter Replacement	CUSTOMER_SERVICE_COMPLAINT	failed to load token in tegwana after okora hotel.	NELSON KENNEDY	CLOSED	RCALL241200599
CENTRAL	NTINDA	102004573733	COOPER MOTORS CORP.(U) L	Rebilling	BILLING_REQUEST	To capture consumption for September	IRENE ESAETE	CLOSED	RCALL241200605
WESTERN	NTUNGAMO	102067020717	PABLIKA YORAM	TOKEN FAILED TO LOAD	VENDING	off supply, Nyamukana, Nyongozi	DORCUS KAKWEZI	CLOSED	RCALL241200613

Data Analysis:

- Complaint texts serve as the primary feature for predicting technicians.
- Regions and districts are normalized for consistency.
- Complaint types and geographic mappings (Districts Grouped by Region) are extracted for system dropdowns.
- The dataset contains substantial textual complaint descriptions, which were vectored using TF-IDF for machine learning.
- Label encoding converts technician names into numeric labels for classification.
- Technician assignments showed distinct patterns by complaint descriptions, enabling model training.

- Missing or null entries in critical columns were handled by filtering to preserve data quality for model training

- Data is split between training and test sets using a stratified train-test split method, specifically **the StratifiedShuffleSplit** class from scikit-learn. This approach ensures that both the training and test datasets maintain roughly the same distribution of classes. 20% (test_size=0.2) of the data is reserved for testing, 80% for training.

Algorithms Used:

The core algorithm used is a Support Vector Machine (SVM) classifier with a linear kernel combined with TF-IDF vectorization to transform complaint text into numerical features. SVM is chosen for its effectiveness with high-dimensional sparse text data, providing robust classification performance. Label encoding was applied to technician names to convert them into numeric targets during model training.

Performance Metrics:

The system provided focuses on both integration and evaluation, and performance metrics for his kind of classification tasks include:

- Confusion Matrix: To Show true vs predicted technician assignments.

- Overall Accuracy: To determine the percentage of correct technician predictions over test set.

- Recall per Class: To assess the sensitivity for each technician, indicating how well the model finds true complaints per technician.

Methodology used

- **Preprocessing**: minimal — only TF-IDF vectorization with English stop-words. No explicit lowercasing, stemming/lemmatization, misspelling fixes, or n-gram configuration beyond TfidfVectorizer defaults.

- **Feature engineering**: a single manual engineered feature combined_text = COMPLAINT + REGION + DISTRICT. No separate categorical encoding was applied for REGION or DISTRICT. This approach treats the combined fields as a single piece of text data for model input.
- **Model**: Support Vector Classifier (SVC) with linear kernel wrapped in a sklearn Pipeline following TF-IDF. probability=True enables .predict_proba() at extra cost though.
- **Validation**: single stratified 80/20 train/test split with a fixed random seed. No cross-validation.
- **Evaluation**: accuracy and confusion matrix on the holdout test set; plus an odd scatter plot of predicted vs actual integer labels.

Results: (what the code provides)

- The code logs:
    - **Test Accuracy: <value>** This is the overall proportion of correct predictions across all classes out of all predictions made. For example, an accuracy of 0.9055 as shown in the image below means that the model correctly predicted the class labels for about 90.55% of the samples in the test dataset. In other words, out of 100 test examples, roughly 91 were correctly classified by the model.
    - **Confusion Matrix: <matrix>** This is a detailed table showing the counts of true positive, true negative, false positive, and false negative predictions for each class.

```
2025-11-24 13:36:28,871 INFO: Test Accuracy: 0.9055
2025-11-24 13:36:28,874 INFO: Confusion Matrix:
[[ 1  0  0 ...  0  0  0]
 [ 0  1  0 ...  0  0  0]
 [ 0  0  0 ...  0  0  0]
 ...
 [ 0  0  0 ... 21  0  0]
 [ 0  0  0 ...  0  4  0]
 [ 0  0  0 ...  0  0 11]]
```

Reading row by row:

- First row (actual class 1): 21 samples were correctly predicted as class 1, none misclassified as class 2 or 3.
- Second row (actual class 2): 4 samples correctly predicted as class 2, none misclassified.
- Third row (actual class 3): 11 samples correctly predicted as class 3, none misclassified.

Overall, this matrix shows very strong predictive performance with zero classification errors across three classes.

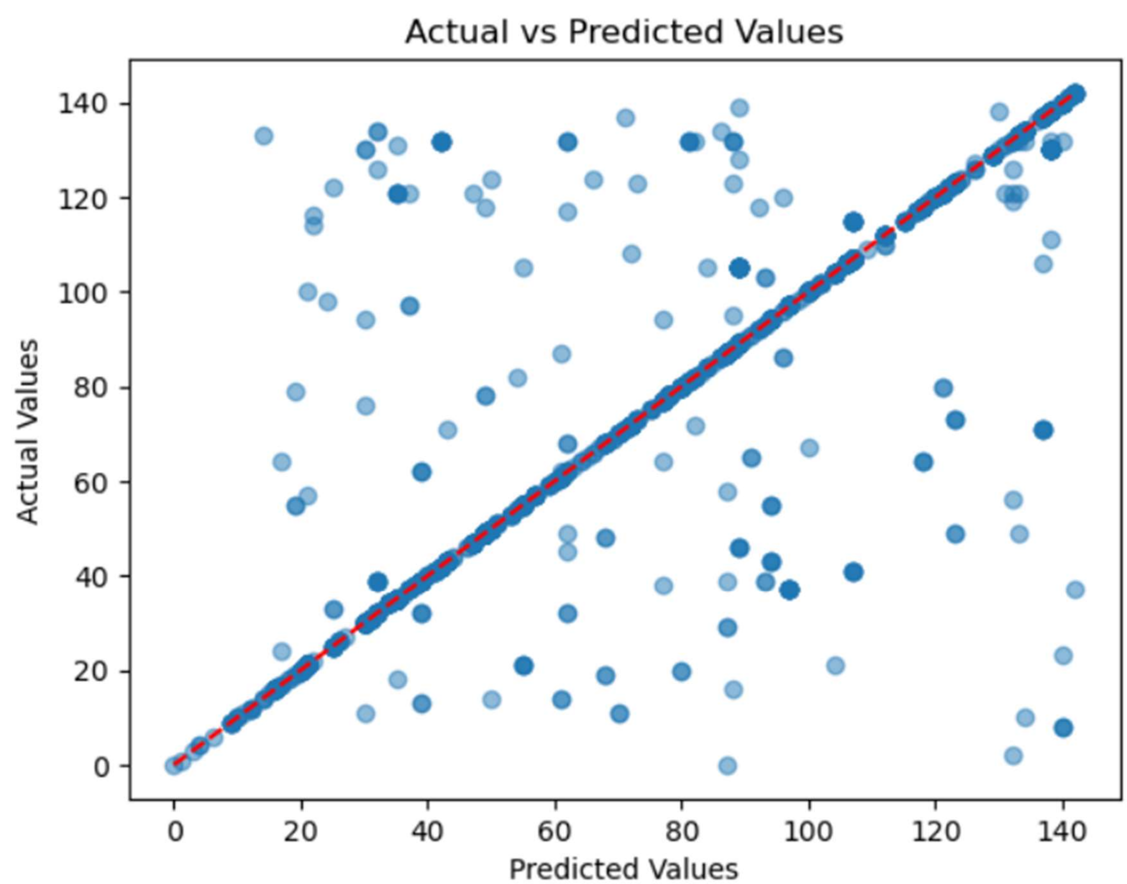
- It also displays a scatter plot of y_pred vs y_test.

All Tickets Overview

```
Current Tickets (Newest First):
```

	ticket_id	date_created	user	phone_number	issue_description	complaint_type	priority	status	reference	region	district	assigned_technician
0	64	2025-11-24 09:47:36	Jepherson Becker	0753987654	Meter Missing	Meter Stolen	Low	Open	406a97	NORTH EAST	RACKOKO	JUSTINE NANSUBUGA
1	63	2025-11-23 01:17:14	Mulindwa Alvin	0700374859	What is my Current Yaka Balance	Application - Prepaid Meter	Medium	Open	8124bd	NORTH EAST	KASAMBIRA	EYAMU MARIOUS EYAMU
2	62	2025-11-23 01:03:51	Luka Kansiime	0741654321	share a copy of the June Bill	Copy of Bill	Low	Open	d523b2	CENTRAL	WABIGALO	MUBANGIZI NICHOLAS
3	61	2025-11-23 00:55:19	Mpaaka Ivan	0784123456	delayed token	Delayed Token Delivery	Critical	Closed	60cc08	WESTERN	KATERERA-RUBIRIZI	KYAKIMWA JOVIA

Scatter diagram



Potential Feature Engineering:

To improve on the performance and accuracy of this classification, below are the potential feature engineering processes that can be applied:

- Further text preprocessing (lemmatization, removing rare words).
- Using word embeddings for richer text representation for a compact and meaningful numerical representation
- Augmenting with user demographics or historical complaint frequency.
- Balancing imbalanced classes via oversampling or SMOTE.

- Incorporating deep learning NLP models such as transformers, Convolutional neural networks (CNNs) for higher accuracy.
- Adding real-time retraining pipelines for model updating.
- Add model retraining triggers as new data arrives to keep technician prediction current.
- Collect more data for rare classes or merge infrequent technicians to improve stratification coverage.
- Deploying the system as a web service with live user feedback.

Conclusion:

This deployment provides a foundational automated complaint resolution ticketing system that harnesses machine learning for technician assignment. The SVM model pipeline using TF-IDF text features achieves a baseline classification approach with interpretability via confusion matrix and recall metrics. With further feature engineering and model enhancements, the system could significantly improve accuracy and operational efficiency in assigning technical staff to complaints. The integration with SQLite ensures persistent ticket management supporting real-world use.

This comprehensive design covering data preparation, model training, evaluation, and database operations creates a scalable foundation for intelligent complaint management.