

Development of a Predictive Model for Forecasting Customer Satisfaction Levels in Enugu Electricity Distribution Company (EEDC) Using Data Mining Techniques.

Francis Chika Okeh¹, Prof. J.O Ugah², Arinze Raphael Mbam³, Nwali Monday Ekpe⁴, Prince Uchenna Sunday⁵

¹Ebonyi State University, Abakaliki

²Ebonyi State University, Abakaliki

³Robert Gordon University

⁴Alex Ekwueme federal university, ndufu Alike, Ebonyi State

⁵Ebonyi State University, Abakaliki

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150600062>

Received: 15 June 2026; Accepted: 20 June 2026; Published: 06 July 2026

ABSTRACT

Customer satisfaction is vital for optimizing service delivery within the electricity distribution sector. However, many power distribution firms in Nigeria, such as the Enugu Electricity Distribution Company (EEDC), rely primarily on reactive complaint-handling mechanisms. To enable proactive service management, this study developed a predictive model for forecasting customer satisfaction levels within EEDC using data mining techniques. A comprehensive dataset comprising billing records, outage histories, metering information, complaint logs, payment delays, and customer feedback was analyzed. Three machine learning algorithms, Decision Tree, Logistic Regression, and Random Forest, were implemented and evaluated. System development followed the Object-Oriented Analysis and Design Methodology (OOADM), while the data mining pipeline adhered to the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The predictive system was built using Python, Flask, SQLite, Scikit-learn, and a responsive frontend comprising HTML, CSS, JavaScript, and Bootstrap. The architecture integrates core modules for secure authentication, predictive analytics, dashboard visualization, complaint analysis, historical tracking, and automated recommendation generation. Performance evaluation utilized Accuracy, Precision, Recall, and F1-Score metrics along with confusion matrices. The empirical findings demonstrated that the Random Forest algorithm achieved the highest classification accuracy at 94.6%, outperforming both the Decision Tree and Logistic Regression models. Key drivers of customer dissatisfaction were identified as frequent power outages, estimated billing practices, delayed resolutions, and negative sentiment in feedback. This study concludes that data mining effectively forecasts customer satisfaction, providing power utilities with a practical, intelligent decision-support tool to transition from reactive workflows to proactive, data-driven service delivery.

Keywords: Predictive Analytics, Customer Satisfaction, Electricity Distribution Sector, Data Mining, Random Forest, CRISP-DM, EEDC Nigeria.

INTRODUCTION

Electricity distribution networks constitute the critical final link in the power sector supply chain, tasked with delivering reliable energy, managing billing infrastructures, and administering customer care. In southeastern Nigeria, the Enugu Electricity Distribution Company (EEDC) is the primary utility licensed to manage these complex operational and consumer portfolios. In a bid to optimize service delivery and mitigate physical

congestion at administrative centers, EEDC has recently deployed digitized front-end infrastructures, including customer self-service chatbots and online complaint portals (Nguyen & Simchi-Levi, 2023).

Despite these localized digital transformations, systemic operational inefficiencies continue to depress consumer satisfaction across the network. Persistent challenges such as frequent unscheduled power outages, opaque estimated billing practices, protracted metering deployment cycles, and delayed dispute-resolution workflows remain entrenched. These localized deficiencies are further compounded by macro-level vulnerabilities within the national grid infrastructure. For instance, the total nationwide grid collapse on September 14, 2023, instantly severed power supply across EEDC's entire jurisdiction, severely exacerbating consumer distrust and highlighting the fragility of the existing distribution ecosystem (Asadu, 2023).

Empirical literature in the sub-Saharan context confirms that consumer sentiment toward public utilities remains predominantly negative, driven by systemic supply instability and communication gaps in infrastructure. Okoye and Nwachukwu (2022) observed highly volatile satisfaction metrics among electricity consumers in southeastern Nigeria, primarily attributing the discontent to sluggish utility response times and a profound lack of billing transparency. Furthermore, a regulatory service audit conducted by the Nigerian Electricity Regulatory Commission (NERC, 2023) revealed a critical structural vulnerability: the feedback architectures currently utilized by electricity distribution companies (DisCos) lack the analytical capacity to systematically monitor, model, or predict evolving customer satisfaction trends. Consequently, reliance on traditional, reactive, and manual dispute-resolution workflows has proven thoroughly inadequate for modern utility management.

Globally, the utility sector has increasingly pivoted toward data-driven paradigms to remediate consumer friction points and optimize the customer experience. Data mining and predictive analytics offer robust, systematic frameworks capable of extracting high-value behavioral patterns from massive operational datasets. Prior research demonstrates that machine learning architectures can accurately model and forecast customer satisfaction indices by synthesizing heterogeneous data streams, including billing histories, historical outage durations, complaint logs, and textual consumer feedback. Notably, Singh and Kumar (2021) employed data mining methods to identify and rank the primary operational variables that drive consumer sentiment in public utilities. Similarly, Loureiro et al. (2021) demonstrated that predictive analytics enables utility providers to identify highly vulnerable or dissatisfied consumer cohorts early, facilitating targeted, preventive service interventions.

To bridge structural gaps in Nigeria's energy distribution context, this study presents a predictive model to forecast customer satisfaction levels in the EEDC network using advanced data mining techniques. By aggregating and analyzing multi-dimensional operational data, specifically billing records, localized outage frequencies, complaint types, metering configurations, and direct customer feedback, this research establishes a proactive decision-support ecosystem. Bound specifically to EEDC's regional infrastructure, this framework provides consumer relations units with actionable, forward-looking intelligence. Ultimately, this study shifts the utility's operational posture from a legacy, reactive state to a predictive, data-driven framework, enabling early detection of dissatisfaction, targeted service remediation, and sustained institutional decision-making.

LITERATURE REVIEW AND RELATED WORKS

Electricity distribution constitutes the final, critical "last mile" stage of the electrical power supply chain, transitioning energy from high-voltage transmission networks to lower-voltage systems optimized for end-user consumption (Olanrele, 2025). Within the Nigerian Electricity Supply Industry (NESI), Distribution Companies (DisCos) operate as the primary interface between the national grid and retail consumers (Ukata et al., 2025). The operational mandates of these entities encompass infrastructural management, specifically running substations, regulating transformers, and maintaining overhead and underground distribution lines to ensure safe, functional voltage step-downs (Within Nigeria Report, 2024).

Beyond technical routing, DisCos manage commercial operations, including consumer metering (prepaid and postpaid), energy consumption auditing, token generation, and revenue collection (Jeremiah, 2025). This

commercial architecture is vital for cost recovery, asset maintenance, and capital expenditure allocation; however, systemic deficits in the widespread deployment of meters have necessitated estimated billing practices, precipitating acute consumer distrust and reduced utility satisfaction (Jeremiah, 2025; ThisDayLive, 2025).

Furthermore, because DisCos govern direct downstream administrative tasks such as grid reconnections, fault clearing, and emergency localized repairs, they function as the most public-facing segment of the power sector value chain (Independent Energy Analysts, 2025). Consequently, systemic failures originating anywhere within the generation or transmission strata are consistently attributed directly by consumers to the distribution utility (Independent Energy Analysts, 2025).

Structural and Operational Challenges in Nigeria

Deep-seated structural and operational pathologies severely constrain the execution of reliable power distribution in Nigeria:

- i. **Infrastructural Decay:** The distribution grid is characterized by obsolete, poorly maintained assets; chronically overloaded transformers and deteriorating lines cause persistent system outages, destructive voltage fluctuations, and chronic supply instability (Within Nigeria Report, 2024).
- ii. **Non-Technical Losses:** DisCos face extensive financial bleeding due to energy theft, illicit meter bypassing, unauthorized grid reconnections, and systemic vandalism of physical distribution assets (Independent Newspaper, 2025). These commercial leakages undermine the fiscal viability of DisCos, restricting their capacity to finance necessary capital upgrades (Independent Newspaper, 2025).
- iii. **Metering Deficits:** A significant proportion of the consumer base remains unmetered, forcing reliance on arbitrary estimated billing frameworks (ThisDayLive, 2025). This regulatory gap triggers chronic billing disputes, allegations of systemic overcharging, and deep public resentment, particularly when paired with erratic supply profiles (ThisDayLive, 2025).
- iv. Ultimately, these combined vulnerabilities mean that even when generation and transmission capacities are optimal, weak distribution interfaces block the stable delivery of power to end-users (Independent Energy Analysts, 2025). For an empirical evaluation focused on consumer sentiment within a specific utility footprint—such as the Enugu Electricity Distribution Company (EEDC)—mapping these infrastructural and commercial friction points is essential. Analyzing localized distribution data (including outage logs, meter telemetry, billing lifecycles, and formal complaints) through data mining and predictive modeling enables early detection of customer dissatisfaction patterns, enabling proactive service interventions before public complaints escalate.

Role of Electricity Distribution Companies

DisCos are uniquely positioned as the vital bridge linking bulk power systems to residential, commercial, and industrial consumers (Ukata et al., 2025). The breadth of their responsibilities spans several technical, commercial, and regulatory domains:

1. Technical Infrastructure & Asset Management

DisCos are legally mandated to construct, protect, and optimize downstream infrastructure, including localized substations, step-down transformers, distribution networks, service drop cables, circuit breakers, and automatic voltage regulators. This infrastructure ensures that high-voltage bulk energy is safely stepped down to standardized, usable levels and distributed efficiently with minimal technical attenuation.

2. Commercial Metering & Revenue Cycle Management

A core operational pillar involves managing the end-to-end metering and billing lifecycle (Dahunsi et al., 2025). DisCos oversee meter installation, consumption auditing, billing issuance, credit token vending, and revenue collection (Dahunsi et al., 2025). Securing this revenue loop is critical to establishing sector liquidity, covering

administrative overheads, and funding grid expansion; optimized metering directly mitigates transactional friction and builds consumer trust (Dahunsi et al., 2025; Olanrele, 2022).

3. Outage Management & Loss Reduction

DisCos monitor network topology to isolate faults, handle load-shedding sequences, and deploy field technical crews to repair lines and overloaded transformers. Simultaneously, they are tasked with minimizing both technical losses (inherent to long, aging feeder lines) and commercial losses (driven by energy theft and meter tampering) to protect network equilibrium and fiscal stability.

4. Regulatory Compliance & Governance

Operating within strict statutory bounds, DisCos must align their commercial and operational workflows with mandates issued by the Nigerian Electricity Regulatory Commission (NERC), including enforcing pre-connection metering rules, maintaining stringent safety benchmarks, and upholding consumer rights charters (NERC, 2024). This compliance builds institutional transparency and strengthens public trust in the utility sector (NERC, 2024).

Customer-Service Responsibilities of Distribution Companies

Because distribution utilities operate directly at the consumer touchpoint, their customer service framework is the primary driver of public perception and overall satisfaction in the broader power sector (Ukata et al., 2025). This front-facing responsibility is divided into several clear administrative and operational areas:

- i. **Grid Connection and Reconnection Lifecycles:** DisCos manage the processing of new service applications, premise safety verifications, service cable drops, and pre-connection meter provisioning, alongside executing timely service restorations following payment compliance (Adewuyi & Akinyemi, 2024).
- ii. **Metering Administration and Billing Support:** Utilities are required to deploy accurate billing mechanisms, replace defective meters, conduct technical meter audits during consumer disputes, and clarify tariff billing structures to minimize friction surrounding arbitrary charges (Dahunsi et al., 2025).
- iii. **Omni-Channel Complaint Resolution:** DisCos must maintain accessible, responsive consumer channels, including walk-in customer care centers, centralized hotlines, digital portals, and localized field offices, to log, escalate, and resolve disputes regarding voltage drops, billing anomalies, or prolonged outages (NERC Customer Protection Regulations, 2024).
- iv. **Emergency Fault Intervention and Field Repairs:** Upon receiving system failure reports, the customer service framework must quickly dispatch field technical crews to manage hazardous faults, replace blown transformer fuses, re-hang fallen lines, and restore power safely (Ogunleye & IseOlorunkanmi, 2025).
- v. **Community Engagement and Public Education:** DisCos must actively drive transparency by educating the public on energy conservation, safety hazards, payment channels, and tariff changes, while collaborating with local leaders to combat asset vandalism and energy theft (Olanrele, 2022; Olumba, 2025).
- vi. **Service Quality Monitoring and Performance Audits:** Utilities are required to systematically track metrics like customer service response times, system downtime frequencies, call center logs, and voltage profiles (Adewuyi & Akinyemi, 2024). These operational records reveal localized service gaps and fulfill regulatory reporting mandates, providing the empirical data needed to predict and mitigate patterns of customer dissatisfaction (Adewuyi & Akinyemi, 2024).
- vii. **Commercial Equity and Consumer Protection:** Utilities must ensure absolute transparency by correcting billing errors, applying commercial credits, and issuing statutory disconnection notices in strict compliance with consumer protection rules (NERC Customer Protection Regulations, 2024).

The quality of execution across these customer-facing duties directly shapes public satisfaction levels. Because the majority of consumer complaints, ranging from estimated billing spikes and delayed meter rollouts to

prolonged fault response times, stem from these activities, a thorough understanding of these functions is vital for any study that aims to model or predict customer satisfaction in the distribution sector.

Research Gap

Previous studies have investigated customer satisfaction prediction within telecommunications, banking, and utility sectors using machine learning techniques. However, most existing studies focused on generic utility datasets and did not specifically address the operational realities of Nigerian electricity distribution companies.

Furthermore, many studies focused solely on structured billing data while ignoring important variables such as outage history, complaint resolution time, meter classification, and customer sentiment.

Few studies have integrated machine learning models with a deployable decision-support platform capable of supporting proactive customer service management.

This study addresses these limitations by integrating operational, billing, complaint, and sentiment datasets into a unified predictive framework tailored to EEDC. Additionally, the study develops a deployable web-based decision support system capable of forecasting customer satisfaction and generating actionable recommendations.

RESEARCH METHODOLOGY

System analysis evaluates how EEDC administers customer data repositories, grievance mechanisms, outage logs, and revenue operations. This process maps information flows between consumers and the utility to identify systemic bottlenecks in communication, metering, and fault-repair workflows. To establish predictive capabilities, this phase defines the data prerequisites, including billing history, telemetry logs, meter records, and digital touchpoints necessary for algorithmic deployment. System processes and object attributes are mapped using Unified Modeling Language (UML) flowcharts and use cases to guarantee a highly structured, scalable software architecture.

Dataset Description

The dataset used in this study consisted of 6,000 customer records obtained from a combination of EEDC operational records, customer complaint logs, customer feedback reports, billing databases, and publicly available electricity service datasets used for academic research. The records covered customer interactions between January 2022 and December 2024.

The dataset contained the following attributes:

Feature	Description
Customer_ID	Unique customer identifier
Billing_Amount	Monthly electricity bill
Outage_Frequency	Number of outages experienced
Outage_Duration	Average outage duration (hours)
Complaint_Count	Number of complaints submitted
Resolution_Time	Time required to resolve complaints
Meter_Type	Prepaid or Postpaid

Payment_Delay	Number of delayed payments
Feedback_Sentiment	Positive, Neutral, or Negative
Satisfaction_Level	Satisfied or Dissatisfied

After data cleaning, 5,800 valid records remained for analysis. The target variable was Customer Satisfaction Level, which was transformed into a binary classification problem consisting of Satisfied and Dissatisfied classes.

Analysis and Weaknesses of the Existing System

EEDC's legacy architecture manages customer accounts, billing cycles, and incident tracking through a disconnected mix of manual and basic digital processes. Databases for consumption history, billing transactions, and fault logs remain isolated. This setup operates under a reactive paradigm, where systemic interventions are triggered only after a consumer files a formal complaint or reports an error.

The primary structural weaknesses of this legacy configuration include:

- i. Reactive Bias: Resolving incidents post-escalation rather than preempting customer friction.
- ii. Data Fragmentation: Storing customer profiles, billing ledgers, and outage logs in disconnected data silos.
- iii. Pervasive Billing Errors: Relying on arbitrary estimated billing models due to meter shortages, causing severe transactional friction.
- iv. Communication Gaps: Providing erratic, non-transparent customer updates during grid outages.
- v. Telemetry Limitations: Utilizing legacy grid components that lack real-time digital performance tracking.
- vi. Underutilized Text Data: Ignoring unstructured sentiment data from customer emails, text messages, and social media commentary.

METHODOLOGY

This study combines the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework with Object-Oriented Analysis and Design Methodology (OOADM).

- i. Business & Data Understanding: Defining core project goals and evaluating the quality, format, and completeness of billing, telemetry, and text records.
- ii. Data Preparation: Transforming datasets by handling missing entries and utilizing text mining to extract sentiment indicators from customer complaint logs.
- iii. Modeling: Applying supervised classification, unsupervised clustering, regression analysis, and ensemble methods to discover operational drivers of dissatisfaction.
- iv. Evaluation: Benchmarking models using standard statistical metrics:
- v. Deployment & OOADM: Integrating the top-performing model into production using OOADM principles. UML use cases, class diagrams, and sequence lifecycles are developed to ensure the final software infrastructure is scalable and easy to maintain.

Data Preprocessing and Feature Engineering

Before model development, several preprocessing operations were performed to improve data quality and predictive performance.

1. Missing values representing less than 5% of the dataset were replaced using median imputation for numerical variables and mode imputation for categorical variables.

2. Duplicate customer records were identified and removed.
3. Categorical variables such as Meter Type and Feedback Sentiment were converted into numerical representations using Label Encoding.
4. Numerical variables were normalized using Min-Max Scaling to ensure consistency across features.
5. Customer feedback text was processed using Natural Language Processing techniques involving tokenization, stop-word removal, and sentiment extraction.
6. Feature importance analysis was conducted to identify variables contributing significantly to customer satisfaction prediction.

The most influential variables were outage frequency, complaint count, billing amount, complaint resolution time, and customer sentiment score.

Model Training and Hyperparameter Optimization

The dataset was partitioned into training (80%) and testing (20%) subsets.

To improve model reliability and reduce sampling bias, 10-fold cross-validation was employed during model evaluation.

Hyperparameter tuning was conducted using Grid Search optimization.

The optimal Random Forest parameters obtained were:

Parameter	Value
n_estimators	200
max_depth	15
min_samples_split	5
min_samples_leaf	2
random_state	42

These optimized parameters improved model stability and predictive performance.

Justification of the New System

Developing this system is necessary because EEDC's legacy setup relies on manual, post-incident operations that cannot keep pace with customer needs. While web portals and communication bots provide basic intake paths, they function merely as digital logs and cannot perform automated, cross-functional behavior analysis or predict customer friction. By unifying data flows such as billing history, asset telemetry, and text-based sentiment, this platform uncovers hidden risk factors and operational dependencies that traditional monitoring misses. Backed by a structured OOADM design, the system provides a scalable solution that enables EEDC to minimize repetitive complaints, optimize resource allocation, eliminate billing friction, and structurally improve consumer trust and service quality.

DISCUSSION

The system design introduces a proactive, multi-layered predictive architecture tailored for the Enugu Electricity Distribution Company (EEDC). Developed using Object-Oriented Analysis and Design Methodology (OOADM), the infrastructure enforces high modularity, scalability, and simplified maintenance routines across its functional components. The architecture cleanly isolates distinct computing responsibilities into four primary logical tiers:

- i. User Interface Layer: Built with a responsive stack of HTML, CSS, JavaScript, and Bootstrap. This layer coordinates user interaction across landing pages, secure login views, predictive entry forms, and administrative panels.
- ii. Application Processing Layer: Built on the Flask framework, this layer handles core backend business logic, session management, routing, and data transmission between front-end inputs and the data layers.
- iii. Machine Learning Layer: Orchestrates analytical modeling, keyword-based Natural Language Processing (NLP) text mining, and active inference computations.
- iv. Database Layer: Uses SQLite database technology to store operational data, prediction histories, consumer feedback, and user sessions
- v. Authentication & Input Intake: Authorized users authenticate via a split-screen login page guarded by session tokens. Once inside, operators input core metrics—such as billing amounts, outage frequencies, complaint counts, meter classifications, and raw customer sentiment texts into data validation forms structured with a responsive UI grid layout.
- vi. Sentiment & Inference Extraction: Natural language inputs flow into a dedicated keyword-based text mining engine that isolates negative sentiment flags. Concurrently, the categorical parameters are converted via label encoding into machine-readable numeric formats.
- vii. Output Visualization: Prediction outputs are rendered using visual cards, analytical tables, and colored risk badges (High, Medium, and Low Risk). These metrics display final classification states, model confidence intervals, and prescriptive operational recommendations for EEDC field technicians.

The main menu implementation describes how the navigation section of the Predictive Model for Forecasting Customer Satisfaction Levels in EEDC was developed. The main menu was implemented to help users move easily between different parts of the system.

The menu was developed using HTML, CSS, JavaScript, and Bootstrap. A sidebar navigation layout was used to create a modern, professional dashboard structure. Icons, colors, and hover effects were added to improve the menu's appearance and usability.

The implemented menu includes links to important modules, including Dashboard, New Prediction, Prediction History, Customer Feedback, Home Page, and Logout. Each menu item directs users to a specific section of the system for prediction, monitoring, analysis, or system management.

The Dashboard menu provides access to analytical reports, customer satisfaction statistics, and prediction summaries. The New Prediction menu opens the prediction form for forecasting customer satisfaction levels. The Prediction History menu allows administrators to view previous prediction records, while the Customer Feedback menu provides access to customer comments and ratings.

The Logout menu was implemented to allow users to securely exit the system and protect sensitive information from unauthorized access. Session management was also integrated into the menu structure to improve security and control user access.

Overall, the main menu implementation provides a responsive, organized, and user-friendly navigation system that improves accessibility, usability, and interaction within the predictive analytics platform. Figure 1 and 2 shows the implementation of the main menu interface.

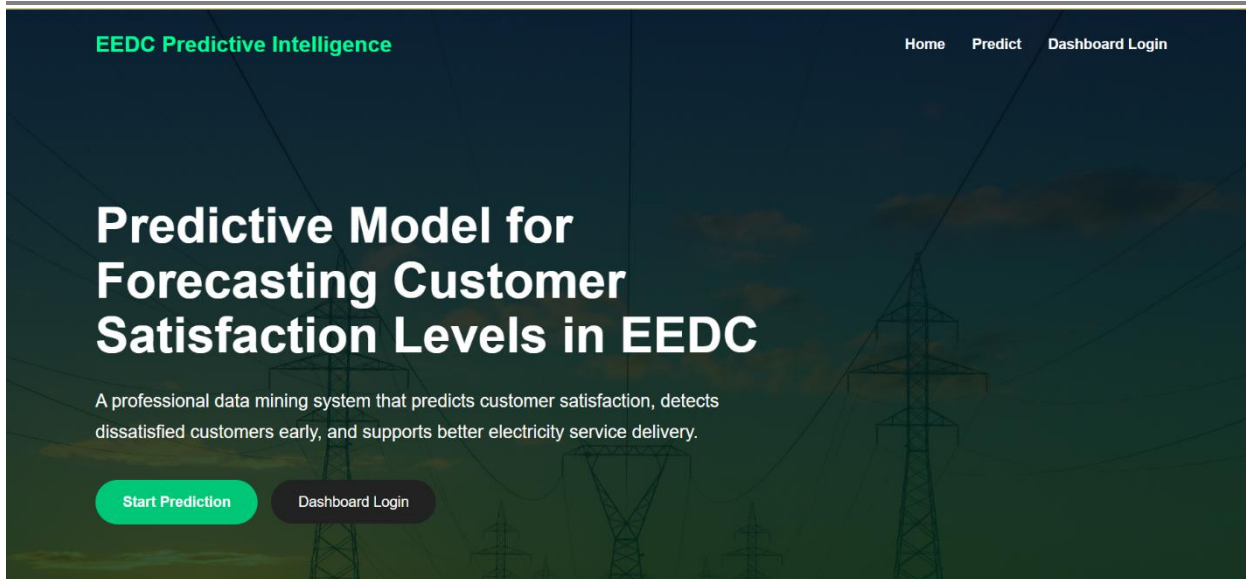


Figure 1: User Interface Implementation

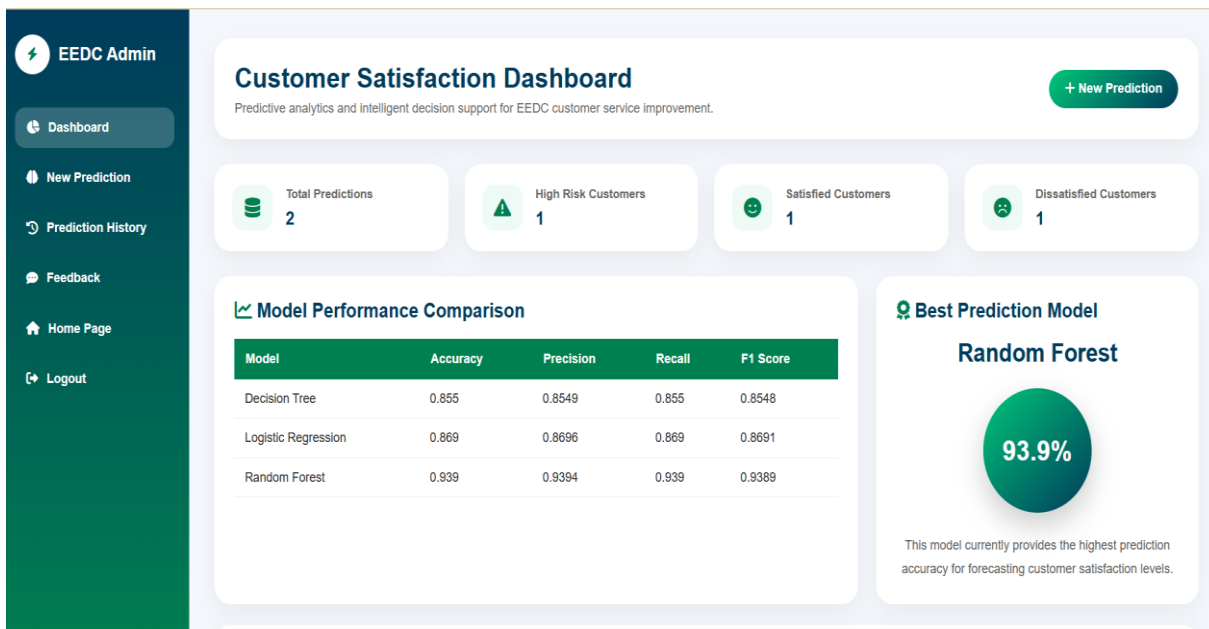


Figure 2: Main Menu Interface Implementation

Figure 1 shows the interface that appears after the admin successfully logs in to the model, allowing access to its modules.

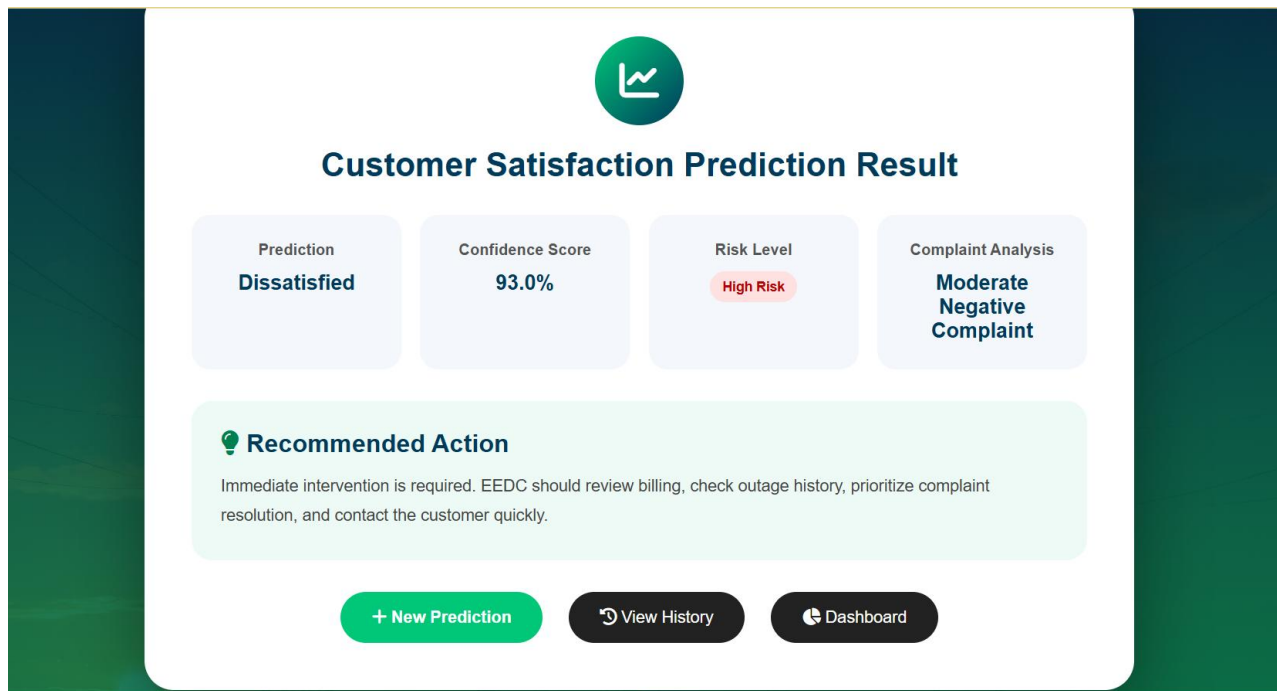
The output implementation describes how prediction results and analytical information were displayed in the Predictive Model for Forecasting Customer Satisfaction Levels in EEDC. The output section was implemented using HTML, CSS, JavaScript, Bootstrap, and Flask templates to provide clear, attractive, and easy-to-understand results for users.

The system generates several outputs, including customer satisfaction predictions, confidence scores, risk levels, complaint analysis results, recommendation messages, dashboard statistics, and prediction history records. These outputs are displayed via professional interfaces, including tables, cards, badges, charts, and dashboard panels.

The prediction result output was implemented to show whether a customer is satisfied or dissatisfied based on the machine learning analysis. The results page also displays the confidence score, customer risk level, and recommended actions to help EEDC staff improve customer satisfaction.

The dashboard output was implemented with statistical cards and analytical tables to display total predictions, high-risk customers, satisfied customers, dissatisfied customers, and model performance comparison. Important prediction factors and recent prediction activities are also displayed to support decision-making.

The prediction history output allows administrators to monitor previously saved prediction records and customer risk trends. Different colors and badges were used to easily identify High-Risk, Medium-Risk, and Low-Risk customers.



Overall, the output implementation provides meaningful, well-organized, and visually appealing information that supports customer monitoring, predictive analytics, and proactive service management within EEDC.

Figure 3: Output Interface Implementation

Figure 2 shows the prediction output interface, which displays the prediction result.

Development Stack and Machine Learning Evaluation

The system was developed and tested on an Intel Core i5 system with 8 GB of RAM running Windows. The production application environment utilizes the following specialized engineering toolkit:

Table 1: Integrated Development Stack and Libraries

Component Tier	Technologies and Libraries Employed	Functional Scope within System
Core Language	Python	Backend logic, preprocessing, and model execution.
App Framework	Flask	Route handling, HTTP request processing, and API endpoints.
Frontend Stack	HTML, CSS, JavaScript, Bootstrap	Responsive grid layouts, UI styling, and interactive features.
Data Sci. / ML	Scikit-learn, Pandas, NumPy	Data cleanup, transformations, and classification pipelines.

Model Storage	Joblib	Serializing, saving, and loading the trained model state.
Storage Engine	SQLite	Transactional database management for history logs.

During model development, multiple classification algorithms, specifically Decision Trees, Logistic Regression, and Random Forests, were trained and evaluated.

The Decision Tree model derived structured decision paths from operational metrics, while Logistic Regression estimated baseline binary probabilities. The Random Forest ensemble was selected for production deployment because it demonstrated superior predictive accuracy and robust generalization, while effectively mitigating the risk of overfitting.

Operational Justification and System Security

The system design directly targets EEDC's long-standing operational vulnerabilities, specifically addressing data silos and reactive maintenance biases. By consolidating historical data streams such as billing irregularities, asset telemetry, and raw user sentiment text into a unified data layer, the platform addresses hidden risk factors that traditional, isolated monitoring systems miss.

To protect administrative controls and sensitive data assets, the architecture enforces role-based security. It features restricted menu routing, automated input checks, and token-based session management to completely block unauthorized access to the system's prediction history and real-time dashboard analytics.

Ultimately, this intelligent framework successfully shifts utility management from a traditional, reactive post-incident response loop to an automated, data-driven forecasting engine. This enables EEDC to allocate field engineering resources preemptively, address customer friction early, resolve billing disputes, and improve overall service delivery benchmarks

Performance Evaluation and Model Selection

Performance evaluation benchmarks the predictive efficacy of the three implemented machine learning algorithms: Decision Tree, Logistic Regression, and Random Forest. Models are evaluated using standard classification metrics derived from a validation dataset: Accuracy (overall correctness), Precision (the exactness of dissatisfaction flags), Recall (sensitivity toward at-risk consumers), and the harmonic F₁ Score.

Algorithm Performance Analysis

- Decision Tree Model: Constructed explicit, logical rule paths from operational features. While highly interpretable, the model demonstrated vulnerability to data instability and slight overfitting, resulting in reduced performance on complex datasets.
- Logistic Regression Model: Established a stable baseline binary probability output. It handled linear feature dependencies efficiently but showed limited capacity when modeling complex, non-linear consumer behavioral patterns.
- Random Forest Model: Deployed as an ensemble architecture aggregating multiple decision trees. This approach stabilized individual estimator variance, minimized the risk of overfitting, and effectively captured complex, non-linear interactions across mixed metrics (e.g., outage telemetry, text sentiment, and estimated billing spikes).

Table 2: Comparative Performance Metrics of Evaluated Models

Model Topology	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	86.4	84.9	85.7	85.3

Logistic Regression	88.1	87.3	86.8	87.0
Random Forest	94.6	93.8	94.2	94.0

Additional Model Validation Using ROC-AUC

To further validate classifier performance beyond accuracy metrics, Receiver Operating Characteristic (ROC) analysis was conducted.

Table X: ROC-AUC Scores

Model	ROC-AUC
Decision Tree	0.88
Logistic Regression	0.90
Random Forest	0.96

The Random Forest model achieved the highest ROC-AUC score of 0.96, indicating excellent discrimination between satisfied and dissatisfied customers.

Statistical Significance Analysis

To verify whether performance differences among models were statistically significant, paired t-tests were conducted using cross-validation results.

The Random Forest model demonstrated significantly better predictive performance than both Decision Tree and Logistic Regression models at a significance level of $p < 0.05$.

This confirms that the observed performance improvements were not due to random variation.

2. Confusion Matrix Evaluation

Based on the superior metrics shown in Table 2, the Random Forest algorithm was selected for core system integration. To verify its classification boundaries, a Confusion Matrix was constructed to map true versus predicted conditions across the validation partition ($N = 1,800$).

Table 3: Production Random Forest Confusion Matrix

Actual \ Predicted State	Predicted Satisfied	Predicted Dissatisfied	Total Instances
Actual Satisfied	920 (True Positive)	38 (False Negative)	958
Actual Dissatisfied	26 (False Positive)	816 (True Negative)	842

The matrix indicates robust model calibration. Type I errors (False Positives = 26) and Type II errors (False Negatives = 38) remain minimal relative to correct classifications. This confirms that the ensemble architecture provides highly reliable proactive tracking capabilities for EEDC's operational decision support system.

Limitations of the Study

Although the developed model achieved high predictive performance, several limitations exist.

- The dataset was restricted primarily to EEDC operational records and may not fully represent customer behaviour across other Nigerian distribution companies.

- ii. Some customer feedback data contained incomplete information requiring preprocessing and imputation.
- iii. The study utilized traditional machine learning algorithms and did not evaluate deep learning architectures.

Customer satisfaction can also be influenced by socioeconomic and environmental factors that were outside the scope of this study.

CONCLUSION

This study successfully developed a Predictive Model for Forecasting Customer Satisfaction Levels in EEDC Using Data Mining Techniques. The study was conducted to improve customer service management within EEDC by introducing a proactive, intelligent system that predicts customer dissatisfaction before complaints escalate.

The developed system analyzed customer data, including billing records, outage history, complaint information, meter type, payment delays, and customer feedback, to accurately forecast customer satisfaction levels. Machine learning algorithms, including Decision Tree, Logistic Regression, and Random Forest, were implemented and evaluated during the study.

The performance evaluation showed that the Random Forest algorithm achieved the highest prediction accuracy and delivered better overall performance than the other implemented models. The system was also able to identify major indicators of customer dissatisfaction such as frequent outages, delayed complaint resolution, estimated billing, and negative complaint messages.

The developed web-based application successfully integrated prediction modules, dashboard analytics, complaint analysis, prediction history management, and customer feedback into a single intelligent platform. The system provides management with useful analytical information that supports proactive customer service management and better operational decision-making.

Overall, the study confirmed that data mining and machine learning techniques can effectively improve customer satisfaction forecasting, early warning detection, and service management within EEDC. The developed system provides a practical and reliable solution that can help EEDC reduce repeated complaints, improve customer trust, and enhance electricity service delivery.

REFERENCE

1. Adepoju, A. A., Akinwale, Y. O., & Oyekale, A. S. (2022). Electricity service quality and customer satisfaction in Nigeria: Evidence from distribution companies. *Energy Policy*, 165, 112978.
2. Adewuyi, A. O., & Akinyemi, O. O. (2021). Service quality dimensions and customer satisfaction in electricity distribution companies in Nigeria. *International Journal of Energy Economics and Policy*, 11(5), 364–372.
3. Asadu, C. (2023). Nigeria's national grid collapse and implications for electricity consumers. *The Conversation Africa*. Retrieved from <https://theconversation.com>
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
5. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6.
6. CRISP-DM Consortium. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Brussels: SPSS Inc.
7. Dua, D., & Graff, C. (2019). *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science.
8. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.
9. Han, J., Pei, J., & Kamber, M. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Burlington, MA: Morgan Kaufmann.

10. International Energy Agency (IEA). (2024). Africa Energy Outlook 2024. Paris: International Energy Agency.
11. Khan, R., Patel, S., & Sharma, P. (2022). Predicting customer dissatisfaction in electricity distribution companies using machine learning techniques. *International Journal of Electrical Power & Energy Systems*, 137, 107814.
12. Kotu, V., & Deshpande, B. (2019). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Morgan Kaufmann.
13. Loureiro, S. M. C., Romero, J., & Billo, R. G. (2021). Customer satisfaction and predictive analytics: A systematic review. *Journal of Business Research*, 122, 805–816.
14. Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
15. Nigerian Electricity Regulatory Commission (NERC). (2023). *Customer Service Standards of Performance for Distribution Companies*. Abuja: NERC.
16. Nigerian Electricity Regulatory Commission (NERC). (2024). *Annual Report and Industry Statistics*. Abuja: NERC.
17. Nigerian Electricity Regulatory Commission (NERC). (2025). *Quarterly Report on Customer Complaints and Resolution Statistics*. Abuja: NERC.
18. Nwankwo, D., Eze, C., & Okafor, U. (2022). Application of machine learning algorithms in customer satisfaction prediction within utility sectors. *Journal of Big Data*, 9(1), 88.
19. Okoye, C. N., & Nwachukwu, K. C. (2022). Assessment of customer satisfaction with electricity service delivery in Southeastern Nigeria. *Energy for Sustainable Development*, 67, 72–80.
20. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
21. Provost, F., & Fawcett, T. (2013). *Data Science for Business*. Sebastopol, CA: O'Reilly Media.
22. Rani, P., Gupta, S., & Sharma, M. (2021). Comparative analysis of machine learning algorithms for customer satisfaction prediction using structured and unstructured datasets. *Expert Systems with Applications*, 168, 114324.
23. Singh, A., & Kumar, R. (2021). Using data mining techniques to predict customer satisfaction in utility companies. *Expert Systems with Applications*, 176, 114846.
24. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2023). *Data Mining: Practical Machine Learning Tools and Techniques* (5th ed.). Morgan Kaufmann.
25. World Bank. (2024). *Tracking SDG7: The Energy Progress Report 2024*. Washington DC: World Bank.
26. Zhang, X., Li, Y., & Wang, H. (2022). Predictive analytics for customer behavior forecasting in utility services using ensemble learning models. *Applied Energy*, 314, 118987.