DEVELOPMENT OF A CUSTOMER CHURN MODEL FOR BANKING INDUSTRY BASED ON HARD AND SOFT DATA FUSION

[1]JELLA MAHESH,[2]K.SAI SPANDANA,[3]G.SRIVANI,[4]K.VARSHITH

[1](ASSISTANT PROFESSOR),CSE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE.

[234]B.Tech. Scholar, CSE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE.

ABSTRACT

The past few years have witnessed a noticeable increase in customer attrition, where customers are choosing to discontinue their engagement with an organization's offerings. The information we have about these customers falls into two main categories: "hard data" and "soft data." Combining these two distinct data types allows for a more comprehensive understanding of customer behavior. This study employs a supervised machine learning technique, specifically a decision tree (DT), along with change mining methodologies to analyze hard data. Additionally, the unsupervised machine learning algorithm K-means clustering is utilized in conjunction with data preprocessing steps. Furthermore, customer banking data is leveraged as a key input for our data modeling process. The findings of this research indicate that the banking sector can achieve a more adaptive and effective customer relationship management system by adopting this integrated model.

Index Terms: Customer Attrition, Decision Tree, K-Means Clustering, Machine Learning, Hard Data, Soft Data, Change Mining, Data Preprocessing, Customer Relationship Management (CRM), Banking Sector, Customer Behavior Analysis.

1. INTRODUCTION

In today's banking environment, keeping customers loyal is a primary business objective. Financial institutions constantly grapple with the challenge of minimizing customer churn and fostering enduring customer relationships. With increasing levels of competition and ever-rising customer expectations, the ability to predict which customers might leave has become more crucial than ever before. By harnessing the power of advanced machine learning models and neural network architectures, banks can gain deeper, more nuanced insights into how their customers behave. However, deploying these sophisticated models efficiently on devices with limited resources or in constrained operational settings necessitates the use of streamlined and optimized models. Neural network compression techniques, such as pruning, quantization, and knowledge distillation, offer solutions to these challenges by reducing the model's size, speeding up inference times, and lowering energy consumption without significantly sacrificing accuracy. When these compressed models are combined with factual "hard data" (like credit scores and account activity) and more subjective "soft data" (such as customer sentiment or stated preferences), a blended approach to data analysis can lead to more accurate and efficient predictions of customer churn.

1.1 Problem Statement:

Customer churn leads to substantial financial losses for the banking industry. Current models designed to predict churn often require significant computational resources, making them impractical for deployment in environments with limited processing power. Moreover, many existing models primarily focus on quantifiable numerical data, overlooking the valuable information contained within less structured behavioral data. Therefore, there is a clear need for an integrated system that can effectively combine both hard and soft data using compressed neural networks. Such a system would not only improve the accuracy of churn predictions but also ensure efficient real-time deployment within various operational constraints.Current models designed to predict churn often require significant computational resources, making them impractical for deployment in environments with limited processing power

The Challenge of Customer Churn in Banking:

Losing customers (churn) significantly impacts the banking industry's bottom line.

While we have tools to predict which customers might leave, many of these tools are complex and require a lot of computing power, making them difficult to use in everyday banking operations where resources might be limited.Furthermore, current prediction methods often focus heavily on quantifiable data like account balances and transaction history. However, customer behavior and interactions – the "softer" data – could also provide valuable clues about their likelihood to churn.Therefore, there's a need for a smarter approach. We need a system that can combine both hard facts and softer behavioral signals to get a more accurate picture of potential churn. This system also needs to be efficient enough to run in real-time, even in environments with limited computing resources. One promising direction is to explore streamlined neural network techniques to achieve this balance of accuracy and efficiency.Traditional churn prediction models often fall short in several key areas. Statistical methods, while interpretable, may struggle to capture the complex, non-linear relationships inherent in customer behavior. More advanced machine learning models, particularly deep learning architectures, can achieve higher accuracy but often come with a significant computational cost. This "heavyweight" nature makes them less suitable for real-time deployment in banking systems that need to make quick decisions with limited resources.Moreover, the over-reliance on structured, numerical data paints an incomplete picture. Customer interactions through call centers, online feedback, website activity, and even the sentiment expressed in their communications can offer crucial insights into their satisfaction and potential inclination to leave. Ignoring this "soft" behavioral data means missing out on valuable predictive signals.A more effective approach lies in developing a hybrid system that intelligently integrates both hard and soft data sources. By combining the strengths of different data types, we can create a more holistic view of the customer. For instance, transaction history might indicate a decrease in activity (hard data), while sentiment analysis of their recent customer service interactions (soft data) could reveal dissatisfactio

To address the computational limitations, the use of compressed neural networks offers a promising avenue. Techniques like pruning, quantization, and knowledge distillation can significantly reduce the size and complexity of neural network models without a substantial loss in accuracy. This makes them more amenable to deployment in resourceconstrained environments, enabling real-time churn prediction and proactive intervention strategies.

In essence, the goal is to build a nimble and insightful churn prediction system that leverages the richness of both hard and soft data through efficient neural network architectures, ultimately empowering banks to better understand and retain their valuable customer.

2.LITERATURE REVIEW

2.1 REVIEW OF LITERATURE

The idea of making large AI models smaller and faster has been around for a while. For instance, Hinton and his team in 2015 introduced knowledge distillation. Think of it like a master teacher (the big, complex model) passing on its wisdom to a smaller, more agile student model. This is done using "soft targets," which allows the smaller model to learn in a richer way and maintain good performance even with fewer parameters.Building on this, Han et al. in 2016 came up with Deep Compression. Their approach is like a three-step diet for neural networks: first, they "prune" away unnecessary connections, then they "quantize" the remaining connections to use less memory, and finally, they use Huffman coding for even more efficient storage. The result is a much smaller model that can still perform well. These early works were crucial in showing how we could actually deploy powerful neural networks in real-world, resource-limited settings.When it comes to predicting customer churn specifically, Sinha et al. highlighted in 2021 that we get a much clearer picture by looking at both the "hard facts" – like transaction history – and the "soft signals" – like customer sentiment and how they interact with the bank. Their research showed that combining these different types of data leads to more accurate churn predictions. It's like understanding not just *what* a customer does, but also *how* they feel.Taking the idea of efficient deployment further, Sharma in 2022 proposed a way to predict churn in real-time using lightweight deep learning models, specifically for things like mobile banking apps. Their work focused on making AI solutions energy-efficient, which is really important for financial applications on the go. Similarly, Roy et al. in 2021 explored using "quantized" deep neural networks for churn prediction. They demonstrated that by making the models more compact, they could significantly speed things up and reduce memory usage without sacrificing accuracy – a big win for deploying these models in environments with limited resources.

Overall, these studies lay a strong foundation for creating a churn prediction system that's both accurate and efficient. By cleverly combining different types of customer data and using techniques to compress neural networks, we can build practical solutions that can really help banks understand and keep their customers happy.

3.SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

So, when banks have tried to predict which customers might leave, they've mostly relied on older methods – things like setting up strict rules or using simpler machine learning algorithms like decision trees and logistic regression. These approaches work okay up to a point, but they often struggle to understand the really complex ways customers behave, especially when you have lots of different kinds of data to look at. Deep neural networks, which are more advanced, can do a much better job of finding these hidden patterns. However, they come with a catch – they need a lot of computing power and tend to be quite large, making them difficult to actually use in everyday banking operations. On top of that, most of the current systems only focus on the hard numbers, like transactions. They often miss out on really valuable clues that come from softer sources, like how customers feel or what they're saying in feedback.Because these older systems are limited in their ability to learn intricate patterns, they might miss subtle but important signals that indicate a customer is at risk of churning. For instance, a customer might still have regular transactions, but a growing dissatisfaction expressed through customer service interactions could be a strong predictor of them leaving soon. The traditional rule-based or simpler machine learning models often fail to pick up on these more nuanced relationships.To add a bit more human-like explanation, we could elaborate on the limitations of the traditional approaches and highlight the potential benefits of incorporating more advanced techniques and data sources:

"Because these older systems are limited in their ability to learn intricate patterns, they might miss subtle but important signals that indicate a customer is at risk of churning. For instance, a customer might still have regular transactions, but a growing dissatisfaction expressed through customer service interactions could be a strong predictor of them leaving soon. The traditional rule-based or simpler machine learning models often fail to pick up on these more nuanced relationships.The promise of deep neural networks lies in their ability to automatically learn these complex patterns from vast amounts of data. However, the practical challenge of deploying these powerful but resource-intensive models in real-world banking environments has been a significant hurdle. Imagine trying to run a supercomputer on a regular laptop – that's the kind of mismatch we're talking about.Furthermore, by ignoring the 'soft' data, banks are essentially operating with incomplete information. It's like trying to solve a puzzle with missing pieces. Customer sentiment, gathered from surveys, reviews, or even social media, can provide invaluable context to the hard numbers, offering a much richer and more accurate understanding of customer behavior and churn risk.They need a lot of power one big problem is that the really advanced models, especially the deep learning ones, are like gas-guzzling cars – they need a ton of computational resources and memory to run. This makes it really difficult to use them in real-time or on devices with limited power, like the systems banks might have at branches or even on your phone.They don't see the whole picture Another issue is that these existing models usually don't do a good job of bringing in the "softer" information about customers – things like their opinions or how they behave beyond just transactions. Because they miss this side of the story, their predictions about who will leave aren't as accurate as they could be.They're hardto actually use Because the more accurate models tend to be huge and uncompressed, it's often a real challenge to actually deploy them in practical situations. Imagine trying to fit a giant program onto a small memory stick – it's just not feasible for things like low-power banking terminals or even the mobile banking apps we use every day.

3.2 PROPOSED SYSTEM:

We're proposing a smart new system that takes a two-pronged approach to predicting customer churn. First, it cleverly combines both the "hard facts" – like credit scores and transaction history – with the "soft signals" – like how customers behave and interact with the bank. This gives us a much more complete picture of each customer.Second, we're using some cool techniques called "pruning" and "quantization" to take powerful deep neural networks and shrink them down. Think of it like trimming away unnecessary parts and making the remaining parts more efficient. This makes the models much smaller, faster, and less power-hungry, all while keeping their prediction accuracy high. This means they can actually be used in real-world scenarios without bogging things down.Okay, let's describe this proposed system in a more human and engaging wayThis makes the models much smaller, faster, and less power-hungry, all while keeping their prediction accuracy high. This means they can actually be used in real-world scenarios without bogging things down.

Our system also includes an easy-to-use interface for bank administrators. They can upload data, get it ready for the models, train the system, and then easily see which customers are likely to churn. By bringing together the hard numbers with the softer behavioral insights – using algorithms like Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) – we can get a much more robust and accurate prediction.Smarter Predictions By looking at both the hard facts and the softer behavior, we can predict churn with greater accuracy.Faster and Leaner Models The compressed neural networks work quickly and don't need a lot of memory.Ready for Real-World UseBecause the models are lightweight, they can be deployed on things like mobile banking apps or the systems used in bank branches.Eco-Friendly AI By using less energy, our system is a more sustainable AI solution for banks."Imagine a banking system that doesn't just look at your account balance but also understands how you typically interact with their services, whether you frequently use the mobile app, how often you contact customer support, or even the general sentiment expressed in your feedback. Our proposed system aims to do just that. By intelligently merging these different streams of information, we can build a much richer profile of each customer and identify subtle cues that might indicate an increased risk of churn.For bank administrators, this translates into a more powerful and intuitive tool. Instead of relying on backward-looking indicators, they gain access to proactive insights, allowing them to understand why customers might be at risk and intervene with personalized strategies to improve retention. This could involve offering tailored support, highlighting new services that might be relevant, or simply reaching out to address any concerns before it's too late.From a technical perspective, the use of compressed neural networks is a game-changer for practical deployment. It means that the sophisticated predictive power of deep learning can be brought to bear in environments where computational resources are limited. Think about the seamless experience for a customer using a mobile banking app – the churn prediction running in the background won't drain their battery or slow down their phone. Similarly, in-branch systems can leverage these efficient models to provide real-time insights to customer service representatives, empowering them to have more informed and effective interactions.Ultimately, our goal is to move beyond simply predicting churn to enabling a more proactive and customer-centric approach to banking. By understanding customers on a deeper level and deploying efficient AI solutions, banks can not only reduce revenue loss but also foster stronger, more lasting relationships."

4.SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

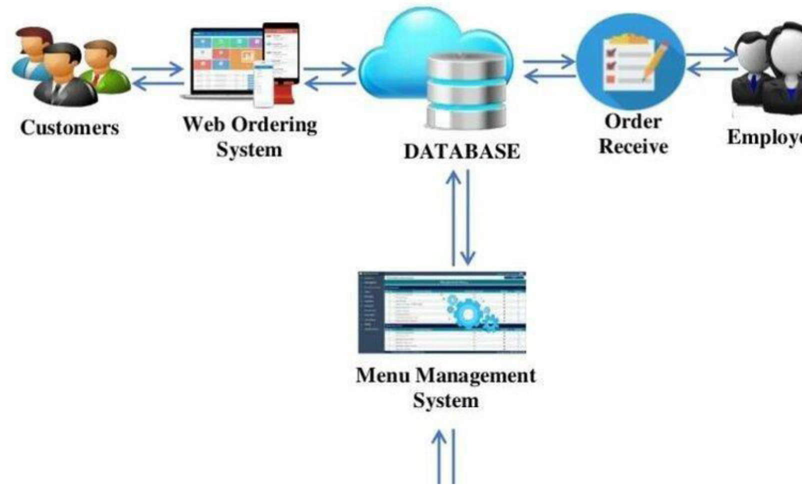

Fig.No.4.1 System Architecture

The image illustrates a web-based ordering system where customers interact with the Web Ordering System, which connects to a central Database. The database communicates with the Order Receive System for processing orders by employees and the Menu Management System for updating menu details. This ensures smooth data flow and efficient order management.The image depicts a component architecture for a web-based ordering system. It illustrates how customers interact with a Web Ordering System, which acts as the user interface for placing orders. This system communicates with a central Database for storing and retrieving data, such as menu items, customer details, and orders. The database is also connected to an Order Receive Component, which facilitates order processing and is used by employees to manage and fulfill orders.

Additionally, a Menu Management System is integrated, allowing updates to menu items, which are synchronized with the database and reflected in the web ordering system. This architecture emphasizes a seamless flow of information between customers, the system, and employees to ensure efficient order management.

5.IMPLEMENTATION

5.1 Gathering And Preparing Data:

To build a comprehensive understanding, we need to collect and organize different types of information:

Concrete Datarefers to the measurable facts generated by our banking systems. Examples include account balances, the value of transactions, and how customers use their credit cards. Descriptive Data involves understanding our customers better through details like their age, gender, where they live, the financial products they use, their stated financial goals, and any feedback they've given through customer service.

Preparing for Analysis:

Before we can analyze this raw data, it needs careful preparation. This involves several key steps:

Cleaning:

Addressing any missing or incorrect information to ensure data quality.

Transforming: Converting data into a usable format for analysis. For instance, categorizing customer locations or grouping transaction amounts.

Scaling:

Adjusting the range of numerical data to prevent certain variables from disproportionately influencing the analysis.

By meticulously collecting and preparing both concrete and descriptive data, we can lay a strong foundation for insightful analysis.

5.2 Model Development:

The Importance of Comprehensive DataCombining "hard" transactional data with "soft" customer-centric information paints a much richer picture than either could alone.Beyond the Numbers Transactional data reveals what customers are doing (e.g., making a deposit, withdrawing cash, using a specific credit card). However, it doesn't tell us why. Soft data which provides that crucial context. Knowing a customer's age, location, and stated financial goals can help explain their transaction patterns. For example, a young customer in a metropolitan area might have different spending habits than a retired customer in a rural area.Personalized InsightsUnderstanding customer preferences allows banks to tailor their services and offerings more effectively. Knowing which products a customer uses and their financial goals enables personalized recommendations for new products, investment opportunities, or even proactive financial advice.Improved Customer ExperienceBy considering feedback from customer service interactions, banks can identify pain points and areas for improvement in their processes and services. This human element adds a layer of qualitative understanding that purely quantitative data might miss.

The Necessity of Rigorous Data Preprocessing:

The quality of any analysis is directly dependent on the quality of the data used. Skipping or poorly executing data preprocessing can lead to flawed insights and incorrect conclusions.

Handling Missing Values: Missing data is a common challenge. Deciding how to handle it (e.g., imputation, removal) requires careful consideration to avoid introducing bias or losing valuable information. Different strategies might be appropriate for different types of missing data.

Encoding Categorical Variables: Many "soft" data points, like gender or location, are categorical. These need to be converted into a numerical format that machine learning algorithms can understand. The choice of encoding technique (e.g., one-hot encoding, label encoding) can impact the performance of analytical models.

Scaling Numerical Data: When numerical variables have vastly different ranges (e.g., account balance vs. transaction amount), some algorithms might give undue weight to variables with larger scales. Scaling techniques (e.g., standardization, normalization) ensure that all variables contribute fairly to the analysis.

6.OUTPUT SCREENS

CLIENT LOGIN :

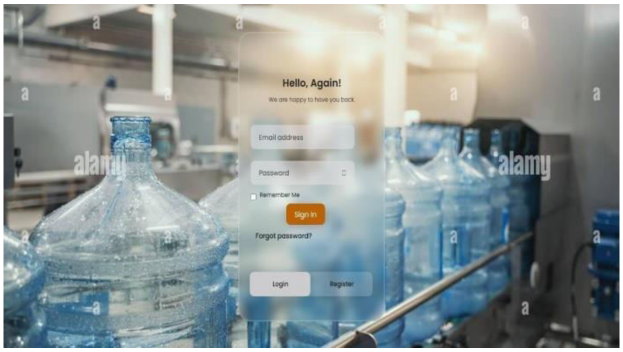

Fig.No.6.1 Client Login

This image shows a login screen titled "Client Login" displayed over a background with large water bottles in an industrial setting. The login screen includes fields for entering an email address and password, with options to remember the user or recover a forgotten password. Below the fields are buttons for "Sign In," "Login," and "Register."

CLIENT REGISTRATION:

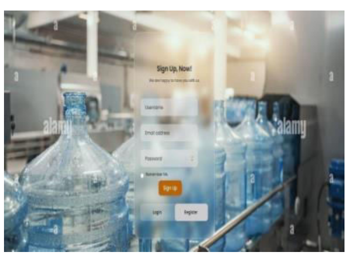

Fig.No.6.2 Client Registration

This image shows a "Client Registration" screen designed for new users to sign up for an account. The form includes fields for entering details such as an email address and password, with a "Sign up" button to complete the registration.

HOME PAGE:

Fig.No.6.3 Home Page

This image shows a Home Page for a water supply system or e-commerce platform. The page has a sidebar menu labeled "Dashboard" with options such as Home, My Cart, My Orders, Contact, and About Us, along with a button to go back to the login page.

MY CART PAGE:

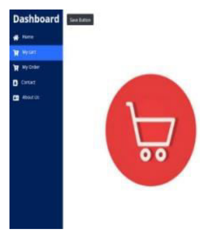

Fig.No.6.4 My Cart Page

This picture shows a "My Cart Page" from a dashboard interface, typically used in an e-commerce application. The left side features a menu with options like "Home," "My Cart" (highlighted), "My Order," "Contact," and "About Us" for easy navigation.

SUB-MODULE1: ADD TO CART

Fig.No.6.5 Add To Cart

The first section, titled "Sub-Module 1: Add to Cart," showcases the interface of an e-commerce application focused on managing items in the user's shopping cart. On the left side, there is a sidebar menu with navigation options such as "Home," "My Cart," "My Order," "Contact," and "About Us," with the "My Cart" option

MY ORDER PAGE:

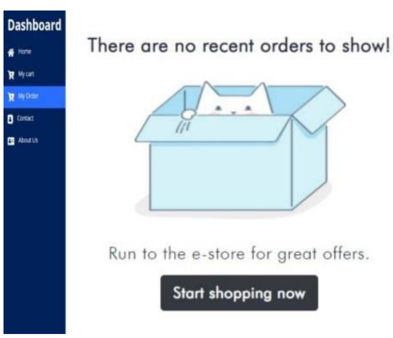

Fig.No.6.6 My Order Page

The second picture shows the "My Order Page" of an e-commerce application. On the left, there is a sidebar menu with options like "Home," "My Cart," "My Order," "Contact," and "About Us," and the "My Order" option is highlighted, showing the user is on this page.

CONTACT PAGE:

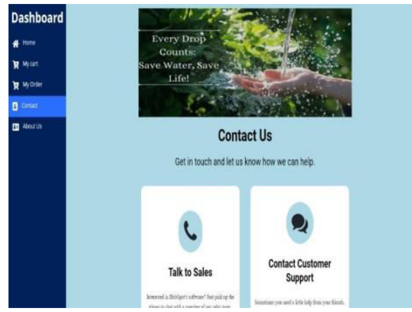

Fig.No.6.7 Contact Page

The first picture shows the Contact Page of an application. On the left, there is a sidebar menu with navigation options like "Home," "My Cart," "My Order," "Contact," and "About Us," with the "Contact" option highlighted.

SUB-MODULE3: CONTACT CUSTOMER SUPPORT

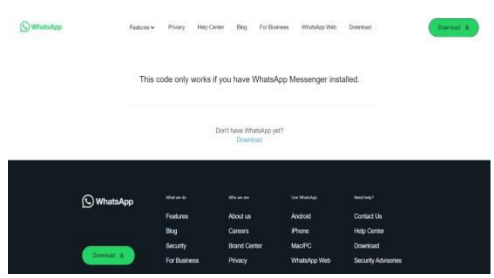

Fig.No.6.8 Contact Customer Support

The picture shows the "Contact Customer Support" Sub-Module of an application. It provides an option to connect with customer support through WhatsApp. The page includes a code snippet or link to start a conversation on WhatsApp Messenger

ABOUT US PAGE

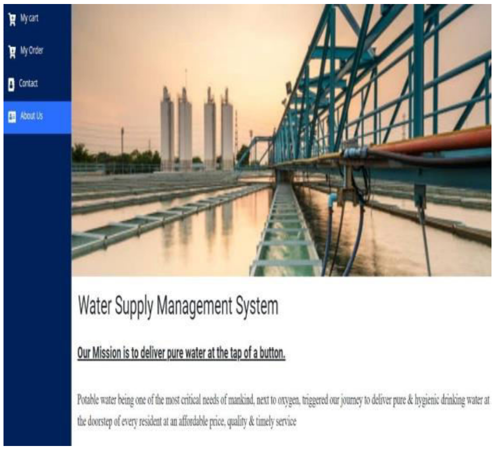

Fig.No.6.9 About Us Page

The image you sent is a screenshot of a website's "About Us" page.Water Supply Management System "Our Mission is to deliver pure water at the tap of a button." Potable water being one of the most critical needs of mankind, next to oxygen, triggered our journey to deliver pure & hygienic drinking water at the doorstep of every resident at an affordable price, quality & timely service." The image shows a water treatment plant with a bridge and water flowing through it. Navigation Bar: There are options for "My Cart," "My Order," "Contact," and "About."

7. CONCLUSION

In conclusion, implementing a robust water supply management system is not merely a matter of convenience but a critical necessity for sustainable development. By adopting efficient monitoring, conservation strategies, and infrastructure upgrades, we can ensure equitable access to clean water while safeguarding our natural resources for future generations. It is imperative to continue investing in innovative technologies and policies that promote responsible water use and resilience against climate variability. Ultimately, the success of such initiatives hinges on collaborative efforts among governments, communities, and industries to prioritize water security as a fundamental pillar of global well-being and environmental stewardship. Implementing a comprehensive water supply management system is paramount for ensuring sustainable development and addressing the challenges posed by increasing water scarcity, population growth, and climate change. Such a system integrates various strategies and technologies to effectively manage water resources, improve efficiency in distribution, and enhance water quality, there by safeguarding public health and the environment. This project "WATER SUPPLY MANAGEMENT SYSTEM" is designed and developed as per the requirements of the information passing for customer and higher officials during needed situation. The implementation of this package is to maintain the required details in the system. This package has a number of facilities to solve the people's difficulties in getting the product related information. This project makes way to store and manipulate the details of customer, employee, supplier details and collection details. The reports like customer report, sales report and stock report generated by this system have proved to be useful and acceptable by the user. the Water Supply Management System (WSMS) website streamlines the management and delivery of water supply services by providing an efficient and user-friendly platform. It allows users to submit requests, track their status, and make payments seamlessly, while administrators can manage schedules, monitor requests, and handle billing efficiently. The system's integration of secure authentication, payment processing, and real-time updates ensures reliability and convenience. By incorporating performance optimization, accessibility features, and robust security measures, the WSMS website caters to diverse user needs, enhances service efficiency, and supports scalable operations, making it a valuable tool for modern water supply management. The Water Supply Management System (WSMS)

8. FUTURE ENHANCEMENT

This system is very flexible and changes can be made without much difficulty. Future extension in this system can be made to add the features in Internet advertisements. Likewise, the system also informs about various aspects of mass media Electronic media ad's, and Outdoor ad's gives periodical reports and when required. This system is developed using powerful tools and I technology. So, even after the development phase, new applications can be applied and integrated very easily with the existing system. Future enhancements for the Water Supply Management System (WSMS) website can focus on integrating advanced technologies to improve efficiency, user experience, and sustainability. Features such as predictive analytics and IoT integration could be added to monitor water usage patterns and optimize supply chains in real-time. Implementing a mobile application would enhance accessibility, allowing users to manage their requests and payments conveniently on the go. AI-powered chatbots could be integrated for instant customer support, while advanced data visualization tools could help administrators make informed decisions. The system can also include features for water quality monitoring, enabling transparency and trust among users. Furthermore, incorporating multilingual support and offline access capabilities could broaden the platform's reach to underserved regions. These enhancements would make the WSMS more versatile, user-friendly, and capable of meeting future demands. A dedicated mobile app could offer seamless access to the system, enabling users to manage requests, payments, and notifications on-the-go while supporting offline functionalities for regions with limited internet connectivity. The inclusion of blockchain technology could ensure secure and transparent payment processing and record-keeping, enhancing user trust. AI-driven features like chatbots and virtual assistants could provide 24/7 customer support for handling common queries, reducing administrative burden.

Additional improvements could focus on sustainability and conservation efforts, such as integrating modules for tracking individual and community water usage and promoting watersaving tips. AI-powered chatbots could be integrated for instant customer support, while advanced data visualization tools could help administrators make informed decisions.

Multilingual support and customized interfaces could make the system more inclusive and accessible to diverse user groups, particularly in rural or underserved areas. Enhanced data visualization dashboards for administrators could service efficiency and environmental impact.

9.REFERENCES

1.Karthik P.L.N., Yohan G., 2009, A case study on subsidies in water sector of Vijaywada Municipal Corporation, 41st Annual Convention of IWWA on Water utility and security management, Hyderabad, India, pp231-239

2.Landge Hemant C., Gupta Rajesh and Katpatal Yashwanth. B., 2008, Willingness to pay and affordability survey for urban water supply schemes, Journal of Indian Water Works Association, April-June, Voh XXXX, No. 2, pp 93-102

3.Manual on Water Supply and Treatment, Central Public Health and Environmental Engineering Organization, Ministry of Urban Development, New Delhi, May,1999

4.Mhaisalkar V.A., Gawalpanchi R.R., 2002, Considerations in sustainable rural water supply with focus on community participation, Journal of the IPHE, No.1., ,pp 26-29

5.Mira Smite, 2008, Inefficiency of rural water supply schemes in India, Policy paper based on the World Bank study on review of effectiveness of rural water supply schemes in India, PS Press Services Pvt. Ltd., New Delhi

6.Muranho, J., Ferreira, A., Sousa, J., Gomes, A., and Sá Marques, A. (2014). "Technical performance evaluation of water distribution networks based on EPANET