**BlockchainAI: WATER LEKAGE**

**GE19612 - PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP PROJECT REPORT**

***Submitted by***

**SHREYA MRIDULA G (2116220701271)**

**SRINIRANJAN V (2116220701287)**

**SURUTHI S (2116220701294)**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

***in***

# COMPUTER SCIENCE AND ENGINEERING



**RAJALAKSHMI ENGINEERING COLLEGE ANNA UNIVERSITY, CHENNAI**

# MAY 2025

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

**BONAFIDE CERTIFICATE**

Certified that this Project titled **“ Blockchain & AI : The Ultimate Sheild Against Fake Identities Online”** is the bonafide work of **“RANJITH KUMARAN G (2116210701209), SAM LAWRANCE V(2116210701225)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

# SIGNATURE

## Dr. P. Kumar., M.E., Ph.D.,

**HEAD OF THE DEPARTMENT**

Professor

Department of Computer Science and Engineering,

Rajalakshmi Engineering College, Chennai - 602 105.

# SIGNATURE

## Dr. M. Rakesh Kumar., M.E., Ph.D.,

## SUPERVISOR

Assistant Professor

Department of Computer Science and Engineering,

Rajalakshmi Engineering College, Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on

**Internal Examiner External Examiner**

# ABSTRACT

# "Blockchain & AI: The Ultimate Shield Against Fake Identities Online" is an advanced system designed to combat the increasing threats of fake social media profiles, misinformation, and identity fraud. By integrating machine learning techniques such as Gradient Boosting, Random Forest, and Support Vector Machine, the system effectively analyzes key profile attributes, including profile picture presence, username structure, description length, external URL usage, account privacy settings, and engagement metrics like the number of posts, followers, and follows. Leveraging ensemble learning methods enhances detection accuracy and reliability This platform is developed as a Flask-based web application, incorporating blockchain technology to ensure data security, transparency, and tamper-proof verification of user profiles. The blockchain ledger maintains an immutable record of profile verification statuses, enhancing user trust and platform integrity. The system's performance is evaluated using precision, recall, and F1 score metrics to optimize fraud detection accuracy. By providing a scalable, secure, and user-friendly solution, "Blockchain & AI: The Ultimate Shield Against Fake Identities Online" aims to create a safer digital ecosystem. It addresses online fraud and misinformation challenges while ensuring the authenticity of user interactions. This innovative approach offers a robust framework for social media administrators and cybersecurity professionals to detect and eliminate fake profiles in real-time, fostering a trustworthy online environment.

# 

# ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides **Dr. JINU SHOPIA.** and **Dr. M. RAKESH KUMAR**, We are very glad to thank our Project Coordinator, **Dr. M. RAKESH KUMAR** AssistantProfessorDepartment of Computer Science and Engineering for his useful tips during our review to build our project.

**SHREYA MRIDULA G** **2116220701271**

**SRINIRANJAN V** **2116220701287**

**SURUTHI S** **2116220701294**

**TABLE OF CONTENTS**

**CHAPTER NO. TITLE PAGE NO.**

|  |  |
| --- | --- |
| **ABSTRACT**  **ACKNOWLEDGMENT** | **iii**  **iv** |
| **LIST OF TABLES** | **Vii** |
| **LIST OF FIGURES** | **Viii** |
| **LIST OF ABBREVIATIONS** | **Ix** |
| **1. INTRODUCTION** | **1** |
| 1.1 GENERAL | 1 |
| 1.2 OBJECTIVES | 2 |
| 1.3 EXISTING SYSTEM | 2 |
| **2. LITERATURE SURVEY** | **3** |
| **3. PROPOSED SYSTEM** | **14** |
| 3.1 GENERAL | 14 |
| 3.2 SYSTEM ARCHITECTURE DIAGRAM | 14 |
| 3.3 DEVELOPMENT ENVIRONMENT | 15 |
| 3.3.1 HARDWARE REQUIREMENTS | 15 |
| 3.3.2 SOFTWARE REQUIREMENTS | 16 |
| 3.4 DESIGN THE ENTIRE SYSTEM | 16 |
| 3.4.1 ACTIVITYY DIAGRAM | 17 |
| 3.4.2 DATA FLOW DIAGRAM | 18 |
| 3.5 STATISTICAL ANALYSIS | 18 |
| **4. MODULE DESCRIPTION** | **20** |

* 1. [SYSTEM ARCHITECTURE20](#_TOC_250016)
     1. [USER INTERFACE DESIGN 20](#_TOC_250015)
     2. BACK END INFRASTRUCTURE [21](#_TOC_250014)

4.2 DATA COLLECTION & PREPROCESSING 21

4.2.1 DATASET & DATA LABELLING 21

* + 1. DATA PREPROCESSING 21

4.2.3 [FEATURE SELECTION 22](#_TOC_250009)

4.2.4 [CLASSIFICATION & MODEL SELECTION 22](#_TOC_250008)

* + 1. [PERFORMANCE EVALUATION 23](#_TOC_250007)
    2. [MODEL DEPLOYMENT 23](#_TOC_250006)

4.2.7 [CENTRALIZED SERVER & DATABASE 23](#_TOC_250005)

4.3 SYSTEM WORKFLOW 23

4.3.1 USER INTERACTION 23

4.3.2 FAKE PROFILE DETECTION 24

4.3.3 BLOCKCHAIN INTEGRATION 24

4.3.4 FRAUD PREVENTION & REPORTING 24

4.3.5 CONTINUOUS LEARNING AND 24

IMPROVEMENT

1. IMPLEMENTATIONS AND RESULTS 24
   1. [IMPLEMENTATION 25](#_TOC_250004)
   2. [OUTPUT SCREENSHOTS 25](#_TOC_250003)
2. CONCLUSION AND FUTURE ENHANCEMENT 30
   1. [CONCLUSION 30](#_TOC_250002)
   2. [FUTURE ENHANCEMENT 30](#_TOC_250001)

[REFERENCES 32](#_TOC_250000)

PUBLICATIONS 36

## LIST OF TABLES

**TABLE NO TITLE PAGE NO**

3.1 HARDWARE REQUIREMENTS 13

3.2 SOFTWARE REQUIREMENTS 14

3.3 COMPARISON OF FEATURES 19

## LIST OF FIGURES

**FIGURE NO TITLE** **PAGE NO**

3.1 SYSTEM ARCHITECTURE 15

3.2 ACTIVITY DIAGRAM 17

3.3 DFD DIAGRAM 18

3.4 COMPARISON GRAPH 19

4.1 SEQUENCE DIAGRAM 20

5.1 DATASET FOR TRAINING 26

5.2 PERFORMANCE EVALUATION AND OPTIMIZATION 27

5.3 CONFUSION MATRIX 27

5.4 BLOCKCHAIN INTEGRATION WITH FLASK FRAMEWORK 28

5.5 WEB PAGE FOR FAKE PROFILE PREDICTION 28

5.6 PREDICTION RESULT 29

**CHAPTER 1 INTRODUCTION**

* 1. **GENERAL**

Proper and secure water distribution is important as an integrative factor for balanced growth in urban or rural settings. Nevertheless, global water distribution systems continue to face challenges like unnoticed leakages, illegal water access via connections and motors, unregulated supply pressure, and fluctuating back pressure. All these factors lead to tremendous wastage of water, reduced supply efficiency, unequal distribution, and breach of water integrity. The traditional methods of monitoring these problems are mostly manual, reactive, outdated, and technologically insufficient concerning the needs of contemporary society. This necessitates the design of an automated, intelligent system capable of monitoring the flow and pressure of water at any given time, identifying irregularities and inefficiencies, and providing secure distribution. Smarten sensors and analytics can be harnessed to develop effective monitoring systems that detect gaps, unauthorized entrances, and pressure discrepancies in the water supply network in real-time and address them before water is misallocated**.**

# 1.2 OBJECTIVE

With the purpose of improving the efficiency, security, and sustainability of water supply infrastructure management, achieve accurate detection of water leakages, illegal siphonages, unauthorized connections, and unattended motorized doexcuting devices through an integrated smart water distribution monitoring system that automates all siphon control activities and provides continuous water pressure monitoring at the end of the distribution series.

# 1.3 EXISTING SYSTEM

The current methods for automatic detection of fake social media profiles extensively use centralized machine learning models, heuristic policies, and human moderation. These methods evaluate user actions, profile completion, activity feeds, and social networks for signs of foul play. Some platforms employ bot detection frameworks and AI models; however, utilizing centralized databases exposes them to various security risks and breaches while violating user privacy. In addition, these systems tend to be tailored for less advanced accounts that employ highly sophisticated impersonation tactics and behavioral mimicry. Lack of scalability and obscurity in dealing with automation means unrestrained manual moderation further undermines their efficiency and effectiveness, enabling the proliferation of issues related to disinformation, scams, and deepening hostility among users.

# CHAPTER 2 LITERATURE SURVEY

**1. Introduction**

Water supply networks are the backbone of contemporary societies and offer access to safe and clean water for home, industry, and agriculture consumption. As a result of growth in population and urbanization, demand for water has grown with unprecedented pressure put on old, in many instances poorly maintained infrastructure. One of the biggest challenges for water utilities at present is the large amount of Non-Revenue Water (NRW)—due to losses through leaks, theft, or meter inaccuracies, it is the amount of water that is produced but not billed to consumers.

As per the International Water Association (IWA), NRW levels can be as high as 40% in certain developing nations. Not only is this loss a huge wastage of a precious resource, but it also translates into financial losses, compromised supply reliability, as well as increased operating expenses. Added to this, illicit water consumption and unauthorized siphoning further aggravate the issue, leading to pressure drops and lowering the level of service offered to valid consumers.

To tackle such problems, an increasing interest has been shown in using cutting-edge technologies like the Internet of Things (IoT), smart meters, machine learning algorithms, and hydraulic modelling. These technologies allow for real-time monitoring, automated detection of anomalies, and data-enabled decision-making in order to improve the management of water supply networks.

This literature review provides a comprehensive overview of the current research and innovations in four key areas: water leakage detection, unauthorized connection detection, illegal water siphoning monitoring, and tail-end pressure surveillance. It also describes the shortcomings of existing systems and formulates the goals of an integrated solution that integrates multiple functionalities for effective water resource management.

**2. Water Leakage Detection**

**2.1 Conventional Methods**

Water leakage has been a serious problem in distribution networks for a long time. Historically, leakage detection was based on manual checks, acoustic logs, and visual audits. Although used today, they are by their nature reactive, time-consuming, and not able to provide real-time insights. Acoustic loggers are less effective in noisy urban centers, and manual methods cannot accurately detect underground or small leaks.

**2.2 Sensor-Based Detection Systems**

* The use of IoT-empowered sensors—particularly pressure and flow sensors—has transformed leakage detection. The deployment of sensors at strategic points within the water network enables real-time monitoring of pressure and flow readings. Deviations from normal behaviour, particularly if pronounced, can indicate the emergence of a leak.
* Wu et al. (2018) created a real-time leakage detection system through the use of IoT devices to measure pressure differential and flow rate variation. This system allowed for rapid response to leaks and water loss reduction. The information was sent to a cloud server where machine learning models processed it to identify patterns related to leaks.
* Romero-Ben et al. (2020) integrated hydraulic simulation with sensor information to forecast leak positions from seen vs. estimated pressure values. Early detection efficiency improved significantly by their model. Nonetheless, sensor positioning, the deployment cost, and the issue of separating honest high usage from genuine leaks confine the usability of such systems.

**2.3 Applications of Machine Learning**

Machine learning has proven to be a valuable weapon in automating the detection of anomalies. Supervised learning models, having been trained on past usage and leakage data, are able to identify new patterns as normal or suggesting a leak. Unsupervised learning techniques such as clustering and autoencoders are also employed where labeled data is limited.

**3. Detection of Unauthorized Connections**

**3.1 Prevalence and Impact**

Illegal connections, or "water theft," are most frequent in poor and densely populated areas where statutory access to water is limited. Since these connections are frequently not monitored at all due to the lack of systematic monitoring and control at the level of distribution, they may escape detection.

**3.2 Pattern Recognition and Billing Analysis**

* Kwak et al. (2019) introduced a machine learning methodology rooted in pattern recognition to identify anomalous consumption patterns that signaled unauthorized access. Through the analysis of consumer billing records, their model identified accounts with sudden spikes or irregularity in consumption behaviour. The system was highly accurate and could be integrated into smart billing software.
* Arregui et al. (2017) targeted District Metered Areas (DMAs)—portions of the water network with well-defined inflow and consumption. Comparing inflow to billed aggregated usage, they detected differences indicating unauthorized use. The validity of this approach relies on high-resolution metering, which might not be available everywhere.

**3.3 Real-Time Monitoring and Tamper Detection**

Smart water meters with tamper detection features and real-time transmission options are another route for curbing unauthorized use. Such meters trigger alarms upon tampering or if consumption exceeds significantly from the past average. Yet, widespread adoption calls for considerable investment and infrastructure support.

**4. Illegal Water Siphoning through Motors**

**4.1 The Issue of Motor-Based Siphoning**

* Illicit siphoning, normally by electric or diesel **motors**, is a special problem in water supply systems. The motors are capable of taking huge amounts of water quickly, usually from public pipes, upsetting the flow balance and leading to pressure losses at downstream locations.
* This activity is common in agricultural and peri-urban regions where water supply is patchy or inadequately controlled. In addition to unequal access, motor-driven siphoning may lead to pipe bursts, contamination threats, and mechanical deterioration of the delivery infrastructure.

**4.2 Detection by Pressure Fluctuations**

* Mitra et al. (2021) studied sudden pressure fluctuations resulting from illicit motor extraction. They created a detection model that employed past pressure history to define a baseline, followed by detection of abrupt and sharp pressure decreases—characteristics of illegal siphoning.
* Chen and Lin (2020) utilized an array of flow sensors and smart valves to monitor abnormal surges in flow. Once they identified siphoning, their system could initiate automatic valve closure to stop further loss. Nonetheless, separation of siphoning from typical high-demand use is still a challenging issue owing to the dynamicity of demand.

**5. Tail-End Pressure Monitoring**

**5.1 Significance of Pressure Monitoring**

* Pressure monitoring at the end of a water supply main is essential for measuring system-wide health. Pressure reduction at such points tends to indicate upstream leakage, extraneous usage, or system inefficiencies in the form of pipe aging and design issues.
* Giustolisi and Savic (2016) showed the application of hydraulic modeling combined with sensor data to model the predicted pressure at different nodes. By comparing real measurements with simulated ones, network anomalies could be rapidly detected.

**5.2 Predictive Analytics and Edge Computing**

* Shang et al. (2022) presented a predictive model that relied on historical sensor readings and water demand predictions to forecast pressure oscillations. By acting ahead of service degradation, this proactive measure allowed utilities to intervene. Edge computing—locating processing near the source, rather than in the cloud—was employed to minimize latency and enhance system responsiveness.
* Edge-based pressure sensors are able to react to anomalies in real time, notifying operators or initiating automated actions such as pressure control or valve opening. These systems are especially valuable in remote and rural areas with poor connectivity.

**6. Integrated Smart Water Monitoring Systems**

**6.1 Unified Platforms for Monitoring**

* A perfect water distribution monitoring system would combine all the features outlined above—leak detection, unauthorized use monitoring, siphoning prevention, and pressure optimization—into one cohesive platform. This involves coordination of several technologies such as IoT devices, cloud and edge computing, machine learning, and visualization tools.
* Li et al. (2019) proposed such a system, a smart water grid that utilized sensors, data loggers, and cloud-based analytics to monitor the entire water network in real time. Not only did their platform detect anomalies but also remote management and control of valves and pumps.
* Patel and Jain (2020) further developed this concept by integrating Geographic Information Systems (GIS), through which water authorities could visualize and map consumption, leakage hotspots, and risk areas. This created a spatial awareness that improved strategic decision-making and resource planning.
  1. **Challenges of Implementation**

Although with promising results, scaling up such combined systems proves to be challenging. These challenges are deployment costs, sensor battery life constraints, data flooding, rural connectivity, etc. Securing data, privacy, and systems resilience against cyber threats is also essential for trust and operation.

**Challenges and Research Gaps**

There are some major gaps in smart water monitoring as follows:

* Integration: Most solutions deal with a particular problem (e.g., leakage detection) as opposed to presenting an end-to-end integrated system.
* Accuracy: Real-time anomaly detection in noisy data with changing environments remains problematic.
* Scalability: Systems are easy to implement in a controlled trial setting but challenging to deploy citywide or in impoverished regions.
* Cost and Energy: Low-cost and low-power sensors are crucial to sustainable deployment but remain limited due to technological restrictions.
* Security: Security of user information and integrity of remote devices must be ensured in a distributed sensor network.
* Next-generation research must attempt to create modular, scalable, and secure architecture that can be applied across different geographic and economic environments.

**Project Objective**

The aim of this project is to conceptualize and create an overall smart water distribution monitoring system that leverages the best of IoT, real-time data analysis, and machine learning to:

* Detect and locate water leakages by detecting anomalies in flow and pressure measurements,
* Detect unauthorized connections by pattern-based consumption analysis,
* Prevent unauthorized water siphoning by detecting sudden decreases in pressure and unauthorized increases in flow,monitoring tail-end pressure continuously to gauge distribution efficiency and detect upstream malfunctions.
* By creating a unified platform capable of real-time detection, alerting, and automated control, the project aims to minimize water loss, enhance system transparency, and ensure equitable distribution of water resources, ultimately contributing to sustainable water infrastructure.

**CHAPTER 3**

**PROPOSED SYSTEM**

* 1. **GENERAL**

**GOAL**

This project's goal is to create and put into place an integrated smart water distribution monitoring system that can:

* Using real-time pressure and flow anomaly detection to find water leaks
* Using usage pattern analysis and flow mismatches to identify unauthorized connections.
* Monitoring pressure drops and abrupt flow surges to stop illicit water siphoning via motors, and
* To guarantee service dependability and early falt detection, pressure levels at the distribution network's tail end are monitored.
* Particularly in areas with aging infrastructure or inadequate water governance, the system seeks to minimize water losses, increase infrastructure efficiency, and encourage fair water distribution.

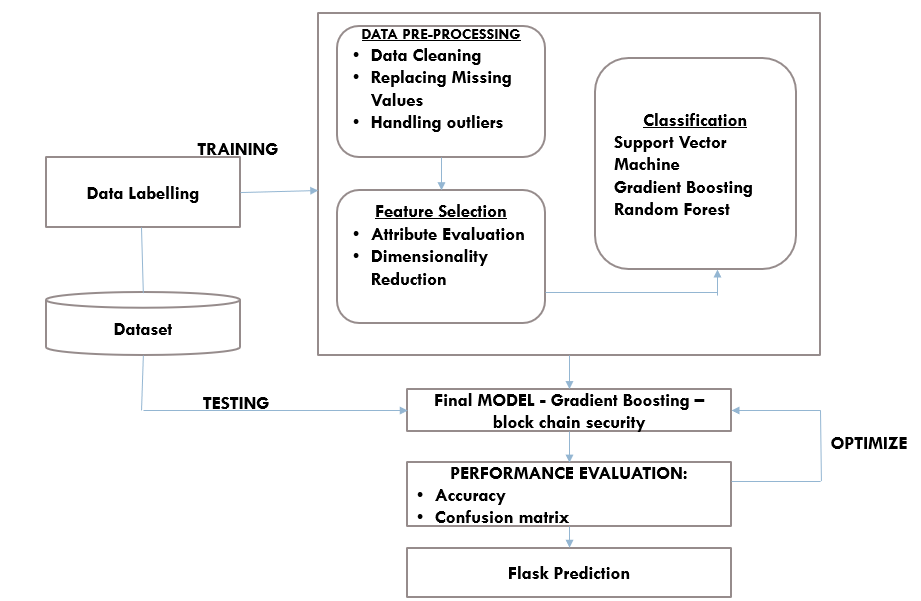
**3.1 PROPOSED GENERAL**

An inventive solution to the urgent problems of leakage, illegal motor-based siphoning, unauthorized connections, and tail-end pressure loss in water distribution networks is the Smart Water Integrity System: A Real-Time IoT and AI-Based Framework for Water Distribution Security and Efficiency. By integrating a network of Internet of Things-based flow and pressure sensors at strategic points along the pipeline, this system makes it possible to continuously monitor the flow of water throughout the distribution system.

The system finds unusual consumption patterns and looks for possible illicit use or leaks using machine learning algorithms like K-Nearest Neighbors, Random Forest, and Support Vector Machines. High accuracy in identifying questionable activities, like illegal tapping or siphoning through external motors, is achieved by anomaly detection models that have been trained on both historical and real-time sensor data.

**SYSTEM ARCHITECTURE DIAGRAM**

The system architecture Fig 3.1 for blockchain security using Gradient Boosting integrates machine learning techniques to ensure robust security by involving user roles such as data providers, analysts, and administrators. It consists of key phases: data collection and labeling, preprocessing (including cleaning, handling missing values, and outliers), feature selection through attribute evaluation and dimensionality reduction, and classification using models like Support Vector Machines, Random Forest, and Gradient Boosting (chosen for its high precision). The performance is evaluated using accuracy metrics and confusion matrices, and the Gradient Boosting model is optimized for real-time anomaly detection. The final model is deployed via a Flask-based system to interact with blockchain networks, ensuring real-time security evaluations. All data, including training results, predictions, and evaluations, is stored in a centralized database, while the server facilitates secure processing and communication. The system's features include reliable data labeling, real-time-blockchain systems.



**Fig 3.1: System Architecture**

# DEVELOPMENTAL ENVIRONMENT

* + 1. **HARDWARE REQUIREMENTS**

The hardware specifications could be used as a basis for a contract for the implementation of the system. This therefore should be a full, full description of the whole system. It is mostly used as a basis for system design by the software engineers.

**Table 3.1 Hardware Requirements**

|  |  |
| --- | --- |
| COMPONENTS | SPECIFICATION |
| Microcontroller | **ESP32 or Arduino Mega (with Wi-Fi/Bluetooth support)** |
| Flow Sensors | **YF-S201 or equivalent water flow sensor** |
| Pressure Sensors | **MPX5010DP or similar (for real-time pressure monitoring)** |
| Smart Valves | **Solenoid Valve (12V, with motor control)** |
| Communication Module | **LoRa / GSM / Wi-Fi Module (for remote data transmission)** |
| Power Supply | **+5V / +12V regulated power supply** |
| Processor (for edge computing) | **Raspberry Pi 4 (Quad-core CPU, 2-4 GB RAM)** |
| Cloud Gateway Device | **Any Linux-based system with internet access** |
| Display Unit (optional) | **LCD/LED display for local alerts and data** |
| Battery Backup | **12V DC battery with charging controller** |
| Enclosure | **Waterproof box (for outdoor sensor units)** |

# SOFTWARE REQUIREMENTS

The software requirements paper contains the system specs. This is a list of things which the system should do, in contrast from the way in which it should do things. The software requirements are used to base the requirements. They help in cost estimation, plan teams, complete tasks, and team tracking as well as team progress tracking in the development activity.

**Table 3.2 Software Requirements**

|  |  |
| --- | --- |
| **SOFTWARE COMPONENTS** | **SPECIFICATION / DESCRIPTION** |
| **Programming Language** | Python, C/C++ (for microcontroller integration) |
| **Development Environment** | Arduino IDE / PlatformIO (for ESP32/Arduino), VS Code (for Python) |
| **Embedded OS (optional)** | MicroPython or FreeRTOS (for ESP32 or Raspberry Pi Pico W) |
| **Server-side Platform** | Flask / FastAPI (for backend and API handling) |
| **Database** | Firebase / MongoDB / MySQL (for sensor data storage) |
| **Cloud Platform** | AWS / Google Cloud / ThingSpeak (for data visualization & analytics) |
| **Machine Learning Libraries** | scikit-learn, TensorFlow Lite, XGBoost |
| **Data Visualization Tools** | Grafana / Power BI / Plotly / Matplotlib |
| **Operating System** | Linux (for Raspberry Pi or cloud host) |
| **Communication Protocols** | MQTT / HTTP / HTTPS / LoRaWAN |
| **Security Framework** | Blockchain APIs (for secure and immutable logging |

# DESIGN OF THE ENTIRE SYSTEM

# 

# ACTIVITY DIAGRAM

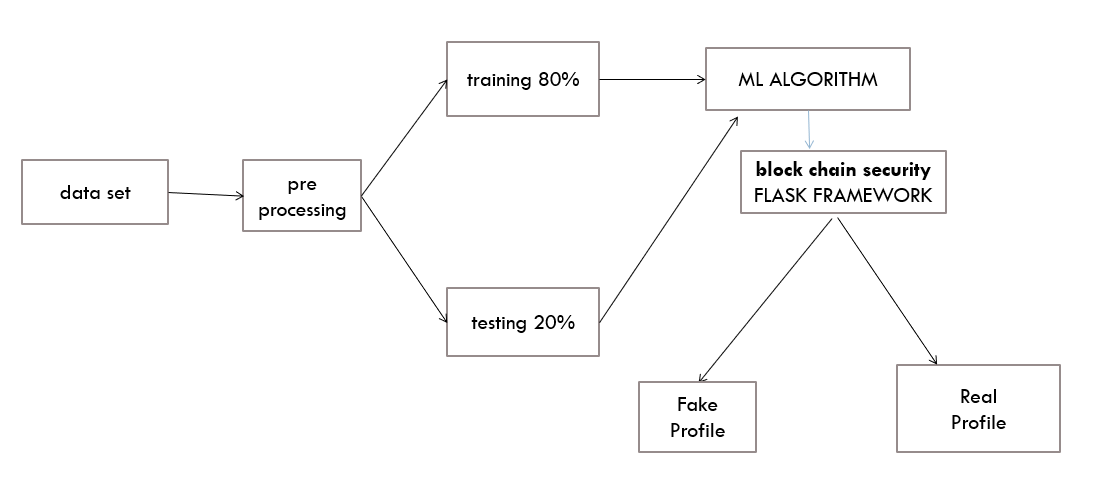
# The activity diagram Fig 3.2 represents the workflow for detecting fake profiles using a Flask-based machine learning system integrated with blockchain security. The process begins with the user interacting via a web page, where they provide the necessary input. The Flask framework serves as the backend, passing the input to a WSGI server for handling requests. The input features submitted by the user, such as profile characteristics, are then sent for preprocessing, where tasks like data cleaning, normalization, and feature extraction are performed. These preprocessed features are passed to the machine learning (ML) algorithm with blockchain security, which processes the data using trained models to classify profiles. The system incorporates blockchain for data integrity and secure operations. Finally, the output, indicating whether the profile is "fake" or "not fake," is delivered back to the user. This streamlined process ensures efficient and secure fake profile detection

# 

**Fig 3.2: Activity Diagram**

# 3.4.2 DATA FLOW DIAGRAM

The data flow diagram Fig 3.3 outlines the process of detecting fake profiles using a machine learning model integrated with blockchain security via a Flask framework. It begins with the dataset, containing raw data on social media profiles, which undergoes preprocessing to handle missing values, remove outliers, and extract relevant features. The preprocessed data is split into training data (80%) for model training and testing data (20%)\* for evaluation. The training phase utilizes machine learning algorithms like Support Vector Machines, Gradient Boosting, or Random Forest. Once trained, the model is deployed with blockchain security and Flask framework for secure, scalable, and tamper-proof operations. The testing phase assesses the model's accuracy, and the system ultimately classifies profiles as either fake or real, ensuring a reliable and secure solution for identifying fraudulent accounts.



**Fig 3.3:Data Flow Diagram**

* 1. **STATISTICAL ANALYSIS**

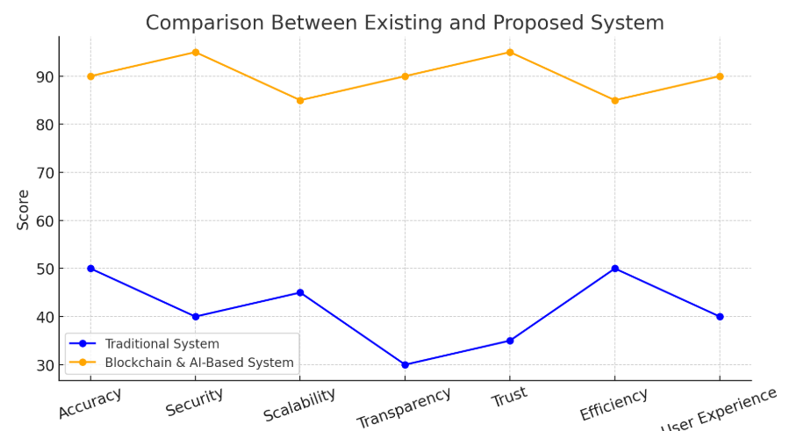
The feature comparison table highlights the key differences between the Blockchain Security Model using Gradient Boosting and traditional blockchain methods. The proposed system integrates advanced machine learning features, including AI-powered anomaly detection, optimized feature selection, and real-time prediction deployment, ensuring a more efficient, data-driven, and secure blockchain environment. While some features overlap with existing systems, the combination of Gradient Boosting and comprehensive optimization enhances threat detection, reduces false positives, and strengthens the overall security of blockchain networks.

**Table 3.3 Comparison of features**

|  |  |  |  |
| --- | --- | --- | --- |
| **Aspect** | **Existing System** | **Proposed System** | **Expected Outcomes** |
| **Threat Detection** | Basic rule-based  anomaly  detection | Al-powered Gradient  Boosting model for anomaly detection | Higher accuracy, reduced false positives |
| **Data**  **Preprocessing** | Minimal data  cleaning and imputation | Comprehensive cleaning, handling missing values and outliers | Improved data quality for training and prediction |
| **Feature Selection** | Limited manual selection | Automated attribute evaluation and dir ionality reduction | Optimized feature set for enhanced model performance |
| **Performance Optimization** | Rarely optimized | Iterative model tuning for Gradient Boosting | Maximized detection capabilities and system  robustness |
| **Deployment** | Manual security evaluation | Flask-based automated prediction system | Real-time, scalable  security evaluations |
| **Scalability** | Limited to specific  Block chain | Adaptable to diverse blockchain applications | Enhanced flexibility and scalability in operations |

The Blockchain & AI-Based System for Fake Profile Detection stands out through its innovative features, distinguishing it from traditional fake profile detection methods. Notably, it integrates advanced machine learning techniques such as Gradient Boosting, Random Forest, and Support Vector Machine to enhance detection accuracy and reliability. Additionally, the system leverages blockchain technology, ensuring data security, transparency, and tamper-proof verification of user identities. Improved scalability and efficiency are key advantages, allowing the system to process vast amounts of data in real time with minimal manual intervention. The decentralized trust model further enhances reliability by eliminating dependence on centralized verification authorities.

Moreover, the platform offers a user-friendly web interface, making interaction intuitive and seamless for users. By significantly reducing false positives, enhancing trust and security, and providing immutable verification records, the system effectively mitigates identity fraud and misinformation. While traditional systems may offer some of these features individually, the holistic approach of the Blockchain & AI-Based System ensures a comprehensive and robust solution for tackling fake profiles. Figure 3.4 depicts the comparative analysis of existing systems versus the proposed system, highlighting its superior performance in various key aspects.



**Fig 3.4 : Comparison Graph**

# CHAPTER 4 MODULE DESCRIPTION

The workflow for the proposed system is designed to ensure a structured and efficient process for detecting and preventing blockchain security threats. It consists of the following sequential steps:

# SYSTEM ARCHITECTURE

# USER INTERFACE DESIGN

# The sequence diagram Fig 4.1 depicts the process of detecting fake profiles, starting with the user providing raw data for data labeling and preprocessing (e.g., cleaning and handling missing values). Features are selected and used to train models like SVM, Gradient Boosting, and Random Forest. The best model is deployed with blockchain security for integrity, tested, optimized, and deployed via Flask for prediction, delivering results to the user.

# 

# Fig 4.1: SEQUENCE DIAGRAM

# BACK END INFRASTRUCTURE

# The backend infrastructure for the sequence diagram comprises a database for managing datasets, storing raw and labeled data for preprocessing, training, and testing. A machine learning framework like TensorFlow or Scikit-learn is used to implement and train models such as SVM, Gradient Boosting, and Random Forest. A blockchain integration layer ensures secure and tamper-proof deployment by storing model metadata and logs, enhancing transparency and trust. Finally, a Flask framework with a WSGI server handles API and application logic, enabling seamless interaction between users and the backend for predictions and result delivery.

# 4.2 DATA COLLECTION AND PREPROCESSING

## 4.2.1 Dataset and Data Labelling

Labeled datasets are collected, including historical blockchain transactions, fraudulent indicators, and legitimate operations. Accurate labeling differentiates between fraudulent and non-fraudulent activities for effective training.

**4.2.2. Data Preprocessing**

The raw dataset undergoes extensive preprocessing, which includes:

Data Cleaning: Elimination of inconsistent or redundant data. Missing Value Replacement: Imputation techniques to handle incomplete entries.

Outlier Detection : Managing extreme or abnormal values for consistency.

**4.2.3 Feature Selection**

Advanced techniques are used to ensure relevant and optimized feature sets:

Attribute Evaluation: Identifying the most influential attributes for threat detection.

Dimensionality Reduction: Reducing data complexity while retaining critical features.

**4.2.4 Classification and Model Selection**

Multiple models are evaluated for classification, such as:

Support Vector Machines: For anomaly detection in blockchain data.

Random Forest: For general-purpose classification tasks.

Gradient Boosting : Selected as the final model for its precision and adaptability in detecting fraudulent activities.

**4.2.5 Performance Evaluation and Optimization**

Model performance is assessed using metrics like accuracy and confusion matrices.

The Gradient Boosting model undergoes iterative optimization to maximize detection accuracy and reduce false positives.

**4.2.6 Model Deployment**

The optimized model is deployed via a Flask-based system, enabling seamless integration with blockchain networks. Real-time security evaluations are conducted by processing live data streams.

**4.2.7 Centralized Server and Database**

All data, including training results, predictions, and evaluations, is stored securely in a centralized database. The server handles communication between the machine learning model and blockchain systems, ensuring secure data processing.

**4.3 SYSTEM WORK FLOW**

**4.3.1 User Interaction:**

Users initiate the verification process by submitting their social media profiles for analysis. The system processes these inputs and evaluates various profile attributes, such as username, profile picture, description, and activity metrics.

**4.3.2 Fake Profile Detection:**

Advanced machine learning techniques (Gradient Boosting, Random Forest, and SVM) are applied to identify patterns associated with fake profiles. The system analyzes multiple factors, including account creation date, username complexity, presence of URLs, privacy settings, and engagement metrics, to determine authenticity.

**4.3.3 Blockchain Integration:**

Once a profile is analyzed, its verification status is securely recorded on the blockchain. This ensures an immutable, transparent ledger that prevents tampering and provides trustworthy proof of authenticity.

**4.3.4 Fraud Prevention & Reporting:**

If a profile is flagged as fraudulent, users receive a detailed report explaining the risk factors. The system allows for further verification steps or appeals, ensuring a \*fair and transparent process. Additionally, the system can automatically alert social media administrators to take action against fake profiles.

**4.3.5 Continuous Learning & Improvement:**

The system continuously updates its machine learning models based on new fraudulent profile patterns. Additionally, user feedback and blockchain records contribute to refining detection accuracy, ensuring that emerging threats are effectively mitigated.

This structured workflow ensures a secure, transparent, and efficient process for detecting and eliminating fake profiles, fostering a safer digital ecosystem.

# CHAPTER 5 IMPLEMENTATION AND RESULTS

# IMPLEMENTATION

The development of the smart water distribution monitoring system integrates IoT, sensor networks, real-time data analytics, and machine learning to identify water leakage, detect unauthorized connections, prevent illegal siphoning, and monitor tail-end pressure in the water distribution network. The system consists of five key modules: sensor deployment, data acquisition, real-time analysis, anomaly detection, and alert/response system.

**5.1 Sensor Deployment and IoT Integration**

* Smart sensors form the backbone of the system. The following sensors are strategically placed in the water distribution network:
* Flow sensors to monitor flow rates at key junctions and endpoints.
* Pressure sensors at various nodes, particularly at the tail-end of pipelines, to detect pressure drops.
* Smart meters installed at consumer endpoints to log water consumption in real time.
* All sensors are connected through an IoT-enabled microcontroller (e.g., ESP32/Arduino + GSM/Wi-Fi modules) to send data to the central system via a secure cloud interface or local gateway.

**5.2 Data Acquisition and Preprocessing**

* Sensor data is collected continuously and sent to a central server or cloud platform (such as AWS IoT Core or Google Cloud IoT). The data includes:
* Flow rates
* Pressure values
* Time-stamped meter readings
* Preprocessing steps include:
* Noise filtering to remove sensor inaccuracies
* Missing data handling
* Timestamp alignment to ensure data synchronization
* Data normalization for machine learning compatibility

**5.3 Real-Time Analytics and Data Storage**

* The backend is built using Python, integrated with a Flask web server and SQLite or Firebase for data storage. Real-time processing is handled using frameworks like Pandas for analysis and MQTT or HTTP REST APIs for sensor communication.
* All incoming data is continuously logged and visualized on a dashboard for monitoring system behaviour and historical trends.

**5.4 Machine Learning for Anomaly Detection**

* The core detection mechanism uses supervised machine learning to identify anomalies:
* Gradient Boosting, Random Forest, and Support Vector Machine (SVM) models are trained using labeled data representing normal and abnormal conditions (leakage, siphoning, unauthorized connections).
* Features include:
* Pressure differentials across nodes
* Sudden flow surges
* Consumption patterns inconsistent with registered usage
* Leakage detection is based on identifying unexpected drops in pressure/flow.
* Unauthorized connections are flagged based on usage behavior anomalies.
* Illegal siphoning is detected from sudden suction-like pressure dips, especially downstream of tail-end sections.

**5.5 Blockchain-Based Data Security**

* To ensure tamper-proof data integrity, the verified anomalies and usage logs are recorded in a blockchain ledger:
* Each record (event or anomaly) is hashed and added as a transaction.
* Ensures transparency and prevents manipulation of leak or fraud reports.
* Integrates seamlessly with the Flask backend for real-time updates.

**5.6 Alert System and Decision Support**

* Upon detection of anomalies, the system:
* Sends real-time alerts via SMS, email, or app notification to the concerned maintenance team.
* Logs the event on the dashboard for auditing and response.
* Optionally triggers automated controls like:
* Valve closure near a leakage point
* Pressure normalization protocols
* Usage cut-off for unauthorized consumers

**5.7 Frontend Interface**

* The frontend dashboard is built using Tailwind CSS and JavaScript frameworks like React or Vue.js. It provides:
* Visualized sensor readings (graphs, heat maps)
* Event logs and alerts
* Map-based view of the water distribution network
* Admin controls for system configuration

# OUTPUT SCREENSHOTS

The project implementation is structured into modules, as depicted in Fig 5.1, highlights the project's seamless integration of machine learning for predictive analysis. It demonstrates a clear workflow, leveraging diverse data inputs for accurate results. The intuitive interface ensures usability across various platforms. Fig 5.2. showcases the project's machine learning model for detecting fake Instagram profiles. It highlights a streamlined workflow, utilizing account metrics for precise predictions. The system ensures adaptability and effective deployment for real-world applications. Fig 5.3 compares the confusion matrices of three classifiers: Gradient Boosting, Random Forest, and Support Vector Machine. It highlights models' performance in distinguishing between fake and non-fake profiles. The visual emphasizes accuracy and misclassification trends, aiding in selecting the best-performing algorithm. Fig 5.4 demonstrates the integration of a machine learning model within a Flask web application, enhanced with blockchain technology for data integrity. The app predicts fake Instagram profiles and securely logs each prediction as a blockchain block. This approach combines predictive analytics with tamper-proof record-keeping for robust and reliable deployment. Fig 5.5 illustrates a Flask web application designed for predicting fake profiles using machine learning. The interface accepts user inputs such as profile picture presence, username characteristics, and privacy settings to assess the authenticity of Instagram profiles. This tool combines user-friendly web design with predictive analytics to provide an accessible and efficient solution for detecting fake accounts. Fig 5.6 presents the prediction result page of the Flask web application. It displays the classification outcome, indicating that the profile is 'Fake,' along with a blockchain-generated hash to ensure the prediction's authenticity and tamper-proof record-keeping. The page includes a 'Go Back' button for navigation, offering a seamless user experience.

# 

Fig 5.1 Dataset for Training

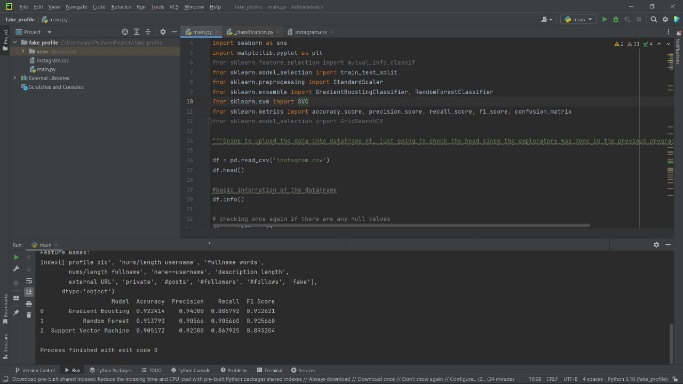


Fig 5.2 Performance Evaluation & Optimization

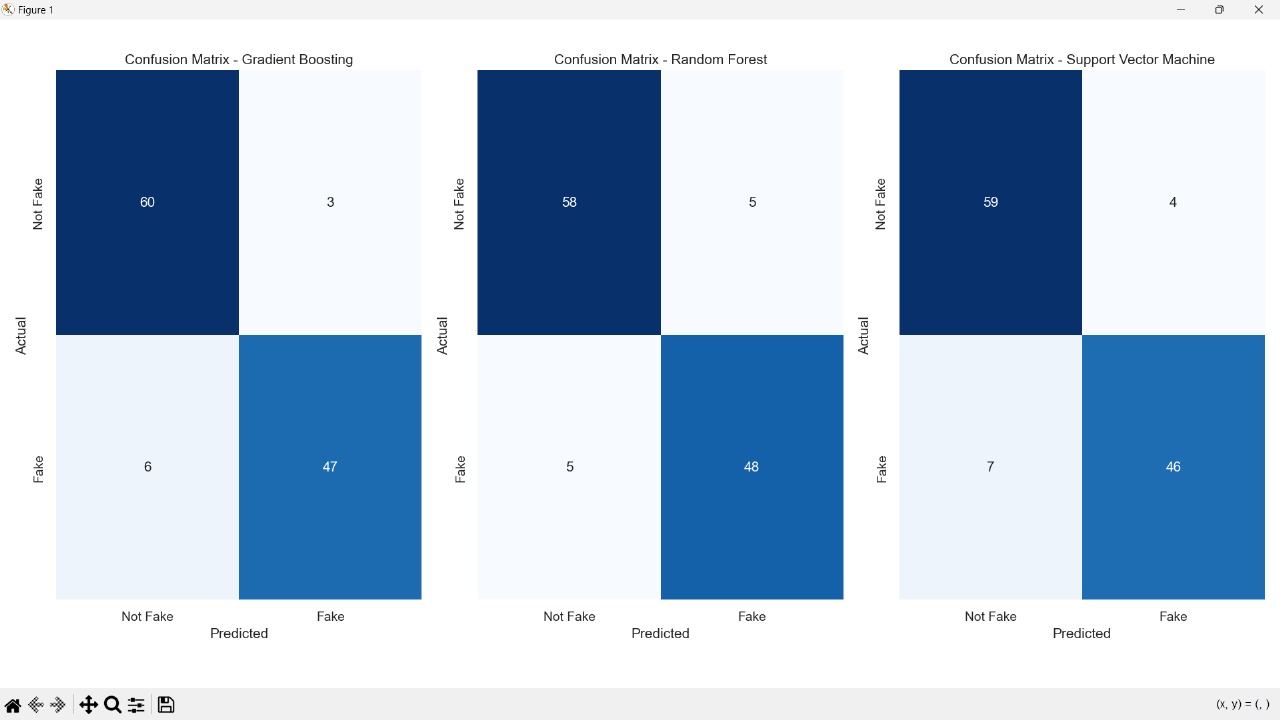


Fig 5.3 Confusion Matrix

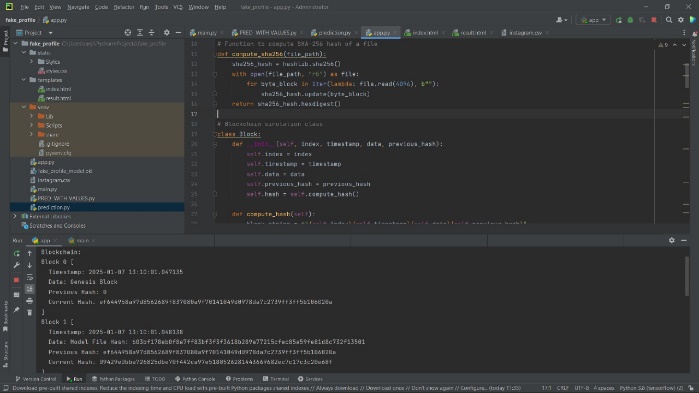


Fig 5.4 Blockchain Integration with flask framework

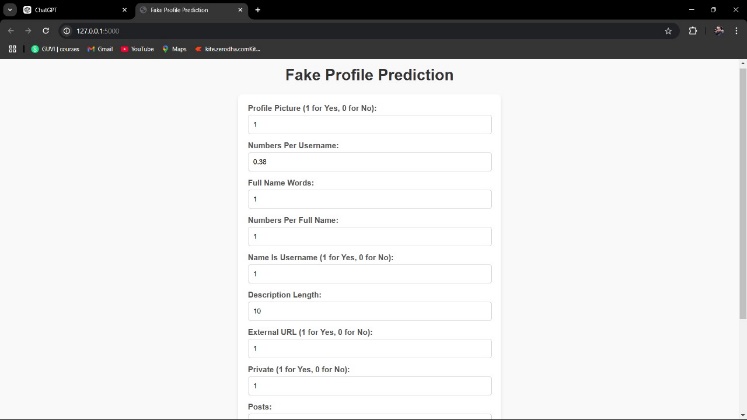


Fig 5.5 Webpage for Fake Profile Prediction

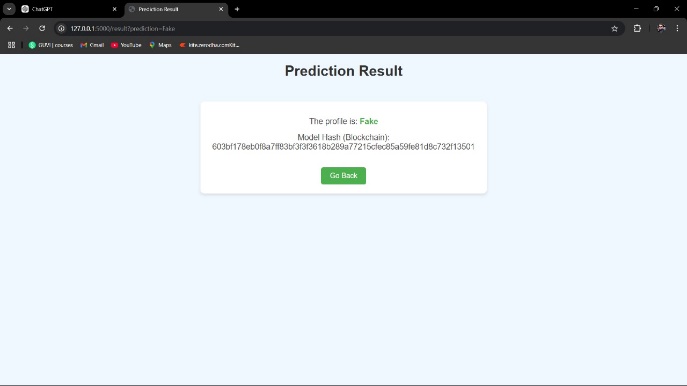


Fig 5.6 Prediction result

**CHAPTER 6**

**CONCLUSION AND FUTURE ENHANCEMENT**

# CONCLUSION

The smart water distribution monitoring system represents a unique turning point in the effective and sustainable management of a valuable resource. With a network of sensor technology combined with IoT technologies and real-time analytics, the system specifically targets large challenges, including the monitoring of water leakage detection, unplanned connections, illegal withdrawal via motors, and low pressure at the tail end of the supply distribution system. It also allows for the use of pressure and flow sensors, together with intelligent data processing, to provide proactive maintenance, curb waste, and improve supply reliability.

This solution optimizes operational efficiency for utilities while promoting equitable access to water distribution, particularly for disadvantaged or vulnerable communities. As water consumption grows globally, intelligent systems, such as these, will be increasingly important for resource management, energy allocation, and accountability of water usage. The system will eventually evolve with enhancements including AI-enabled predictive maintenance, remote/mobile access, and blockchain-enabled security, thereby offering a vast, integrated, scalable platform for smarter, safer, and sustainable urban living.

# FUTURE ENHANCEMENT

* **AI-Enabled Predictive Maintenance Integration**

Employ deep learning models to forecast potential failures or leaks before they happen using long-term historical data, pressure variations, and usage trends data.

* **GIS and Satellite Integration**

Utilize GIS and satellite imagery to identify illegal connections to the pipeline, and follow the distribution of water in cities as it spans far geographical areas.

* **Mobile App for consumers and field technicians**

Create an app that would permit the utility staff to receive alerts in real time, troubleshoot, and track preventative maintenance activities. Consumers will also be able to access the app to monitor their usage patterns and be alerted to anomalies.

* **Solar-Powered IoT Nodes**

Use solar-powered units for sustainability and reduce required maintenance on sensor units that are battery-powered, particularly in rural and remote areas.

* **Edge Computing Integration**

Run small machine learning models on microcontrollers or edge devices to improve the response time of transferring data and facilitate real-time decisions.

* **Smart Billing System**

Link water use data to dynamic billing systems to help create incentives for conservation, and provide alerts for suspicious use so it may be inspected.

* **Self-Healing Pipe Networks (Research)**

Investigation into smart materials, or actuators, in pipes that could recognize and autonomously seal small leaks.

* **Citizens Reporting Platform**

Extend opportunities for the community to report issues concerning the water utility via electronic or mobile means.

# REFERENCES BOOKS

1)"Water Supply and Sewerage" by E.W. Steel and Terence J.McGhee

2)" Water Distribution System Handbook" by Larry Mays

3)"Smart Water Technologies and Techniques" by David A. Lloyd Owen

4)"Introduction to Infrastructure: An Introduction to Civil and Environmental Engineering" by Michael R. Penn and Philip J. Parker

5)"Internet of Things: Architecture and Applications" by Rajkumar Buyya, Amir Vahid Dastjerdi

6)"Hydraulic Modeling and GIS" by Pramod Pandey and Santosh Pathak

7)"Data-Driven Modeling for Sustainable Engineering" by Kondo H. Adjallah

8)"Machine Learning: A Probabilistic Perspective" by Kevin P. Murphy

9)"Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili

10)"Water Resources Engineering" by Larry W. Mays

11)"Handbook of Water and Wastewater Systems Protection" by Robert M. Clark and Simon Hakim

12)"Hydraulics of Pipeline Systems" by Bruce E. Larock, Roland W. Jeppson, and Gary Z. Watters

13)"IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things" by David Hanes et al.

14)"Real-Time Systems Design and Analysis" by Phillip A. Laplante

15)"An Introduction to Water Distribution Systems" by P. H. R. Feddes

16)"Fundamentals of Water System Design" by Robert Pitt

17)"Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and Implementations" by Fei Hu

18)"Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia" by Anthony M. Townsend

19)"Water Resources Systems Planning and Management" by Daniel P. Loucks and Eelco van Beek

20)"Applied Hydraulic Transients" by M. Hanif Chaudhry