Project and Professionalism

(6CS007)

EnVIIsage

(A Healthcare Data Analytics Platform)

Student ID : 2226684

Student Name : Kaushal Rai

Group : L6CG4

Supervisor : Ganesh Kuikel

Reader : Sunita Parajuli

Acknowledgment

My heartfelt gratitude goes to our respected supervisor, Mr. Ganesh Kuikel, and our reader, Ms. Sunita Parajuli, for their constant advice and encouragement throughout the preliminary stages of my initial final year project proposal, which is about "Healthcare Data Analytics Platform." Their knowledge and assistance have been critical in determining how this project has progressed. I am grateful for all of the guidance that has characterized this early stage, and I am eager to apply this knowledge and understanding to the successful completion of the "Healthcare Data Analytics Platform" project.

With due respect,

Kaushal Rai

Table of Content

I. Introduction

As a result of the advent of data analytics, the healthcare sector has experienced a major paradigm shift. This integration allows for evidence-based decision-making, better patient care, and better resource management. By employing cutting-edge technology and methods, the goal of this project is to provide a comprehensive platform for healthcare data analytics.

This platform will allow healthcare professionals to gain useful insights from the massive amounts of healthcare data at their disposal. This data can help with decision-making and improve patient outcomes by employing trend analysis and predictive modeling. The Healthcare Data Analytics Platform will consider population health patterns in addition to individual patient data. Larger-scale data collecting and analysis supports the development of public health campaigns and preventive actions to address prevalent health conditions. Interventions and policies aiming at promoting the overall well-being of the community will be easier to develop with this macro-level view.

II. Project Details

A. Academic Questions

As we continue exploring the relationship between these two sectors, we're curious to find out how technological platforms, artificial intelligence (AI), and data analytics may truly benefit healthcare. These subjects revolve on the use of AI to predict health issues, the use of data to enhance healthcare, and the promotion of informed, evidence-based decision-making through digital platforms.

- How might data analytics lead to better health outcomes?
- How might AI help anticipate and avoid health problems?
- What role does a data analytics platform play in supporting evidence-based procedures?

B. Aims

Our project is driven by goals centered around utilizing technology to transform healthcare. These goals are intended to help you understand how our project hopes to significantly change the healthcare environment by implementing predictive AI models, improving healthcare decision-making, increasing patient engagement, and developing a user-friendly data analytics platform.

- Create an advanced, patient-friendly platform for advanced healthcare data analytics that is accessible to both medical professionals and patients
- Boost proactive patient management and engagement
- Apply AI models to forecast health trends and risks
- Boost healthcare decision-making with data-driven insights.

C. Objectives

- Develop a user-friendly web app for healthcare data analytics
- Perform predictive analysis using AI algorithms
- Use natural language processing to get information out of unstructured health data
- Implement retrieval mechanisms and robust data storage

D. Problem Statement

- The enormous volume of healthcare data makes it difficult to effectively analyze and forecast health trends.
- The limitations of conventional techniques that prevent prompt treatments
- Possible repercussions of postponing interventions as a result of ineffective data processing

E. The Project as a Solution

- The platform is envisioned as a potent resource for medical practitioners to examine patterns, forecast results, and make sound decisions
- Personalized care and proactive health management can be given to patients by anticipating potential health issues and recognizing early warning indicators
- AI integration enables proactive health management and tailored treatment

F. Scope and Limitation

- Create possible partnerships with healthcare organizations to increase the platform's influence and enable a group approach to data-driven decision-making and healthcare analytics
- Appreciate the limitations of real-time data streaming and stress that real-time capabilities might be added in subsequent project phases based on user requests and technological advancements
- Acknowledge that the accuracy and quality of the input data determine how effective the platform is, and stress the significance of continuous data quality assurance procedures.

III. Literature Review

It's a platform for healthcare data analytics. With the help of this platform, medical professionals will be able to make meaningful insights out of the vast volumes of healthcare data at their disposal. By using trend analysis and predictive modeling, this data can aid in decision-making and enhance patient outcomes. We're interested in learning more about how technology platforms, artificial intelligence (AI), and data analytics may improve healthcare as we investigate the connection between these two industries. These topics center on leveraging AI to anticipate health problems, leveraging data to improve healthcare, and leveraging digital platforms to encourage well-informed,

evidence-based decision-making. It assists clinical decision-making in the medical sector by identifying the optimal course of action for the community's betterment through the collection and analysis of community health data.

1. Research Work

The application that is currently being developed is a web-based platform for healthcare data especially for usage by hospitals and organizations. Healthcare personnel working in these environments will be able to effectively examine and understand medical data thanks to this customized design. The web browser accessibility of the platform facilitates its smooth integration into hospital and organizational workflows, enabling workers to obtain important insights from healthcare data across various locations and devices. The purpose of this focused strategy is to improve data-driven decision-making in the context of organizations and healthcare.

2. End User

In utilizing healthcare data analytics platforms, organizations and hospitals stand to gain transformative insights and operational enhancements. This platform supports strategic planning and resource allocation for healthcare decision-makers by providing real-time data integration and predictive analytics. Organizations benefit from reduced administrative procedures and optimal resource allocation, while hospitals can improve clinical outcomes and patient care through customized treatment regimens. While organizations benefit from reduced administrative procedures and optimal resource allocation, hospitals can improve patient care through personalized treatment plans and improved clinical outcomes.

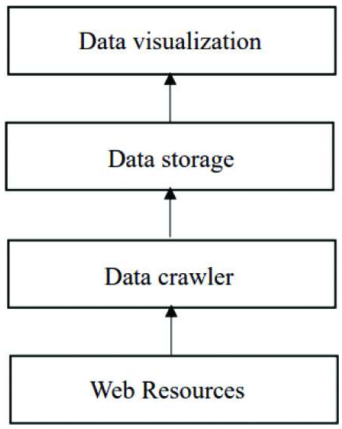
3. Solution

To assist with clinical decision-making, this platform provides features such as a summary of the reports and statistical tools for processing, analyzing, and predicting various outcomes. It allows several users to work together on data analysis projects by providing shared workspaces. These solutions give healthcare decision-makers real-time data integration and predictive analytics to support their strategy planning and budget allocation. This system highlights the crucial role of analytics platforms in population health management, enabling organizations to identify health trends, allocate resources efficiently, and implement preventive measures.

4. Similar Systems

Design and Implementation of Big Data Crawling and Visualization System Based on COVID-19 Data

This system uses Python crawler technology to access COVID-19 data and is based on the Scrapy framework. The cleaned data is then put into a MYSQL database. Django is used to create the front-end interface, while the React environment provides the front-end foundation for the epidemic data. The COVID-19 webpage uses Ant Design technology to combine maps and visual images to provide a versatile and efficient way to visualize pandemic data. The two components of the system are the processing and crawling of epidemic data and the data visualization and analysis.



The system gathers real-time data about epidemics both domestically and internationally using data crawler technology. The information is sourced from national public data sources. The information needed for any epidemic visualization system is retrieved and entered into the database via the appropriate URL data interface. After being cleaned, the crawled data is converted into a JSON data dictionary. The data that was crawled is divided into two sets: national epidemic trend data and worldwide epidemic dispersion data, following cleaning, JSON data conversion, and other procedures. Extracting, classifying, and storing the global cumulative and real-time epidemic data in the database under the "global_history_URL". Processing, saving, and traversing data under the database table structure are all part of the process. Launch the relevant request under the data specifications, then wait for the server to answer. A data set is created by processing the returned data. The data collection is finally visualized. This is essential for raising knowledge of self-defense techniques and stopping the epidemic from spreading.

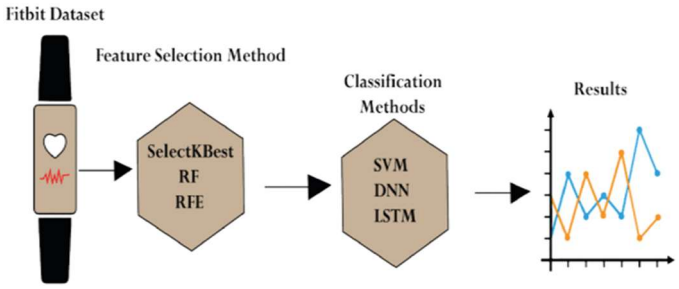
Machine Learning Technique to Monitor Heartbeat Using Amalgamated Data of Multi-Sensor Stream

Introduction

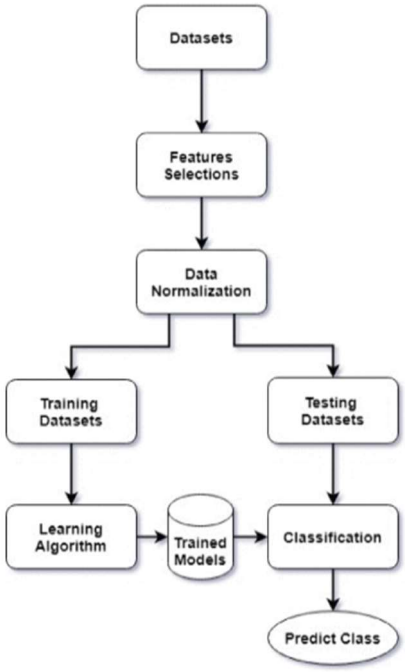
The present research utilizes machine learning algorithms, including LSTM, DNN, and SVM, to monitor chronic conditions, such as heart rate, using wearable devices. The objective is to use Random Forest to predict abnormal behavior patterns in FitBit data with an accuracy rate of 85%. The research offers fitness behavior guidance and can help predict and avoid diseases.

Methodology

The Human Body Sensor Activity Network (HBSAN) is the proposed technique for this research. It will gather combined sensor data, identify everyday activities, and predict unanticipated diseases.



The dataset is assigned to the training and testing phase to learn the model features and provide real-time problems. The sensors' data first go through the feature selection phase, following which it goes through the data normalization phase where data labeling and data scaling are done. The dataset is split, the training dataset is entered into the learning algorithm to train the model, and the testing dataset is sent into the classification process to predict and make decisions about the input, as seen in the figure below.

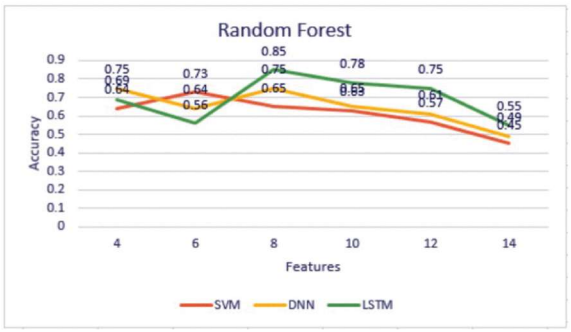


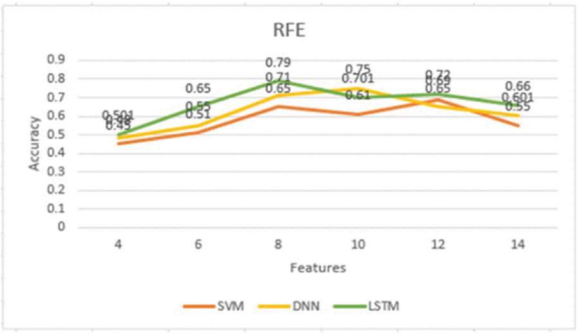
Results

The dataset, which was split into 70% of the training dataset and 30% of the testing dataset to analyze the performance variation, was used to select the best features and

provide inputs to SVM, DNN, and LSTM for the classification process. Three classifiers—the univariate method, RFE, and RF—were used for feature selection. Based on the FitBit dataset, RF produced the best results in the analysis as compared to other methods.

Feature Selection	SVM	DNN	LSTM
SelectKBest	70.44	77.64	79.54
RFE	72.99	80.75	81.63
Random Forest	75.25	82.04	85.05





Conclusion

Wearable technology is essential for healthcare monitoring, especially fitness trackers like Fitbit. Using SVM, DNN, and LSTM models in conjunction with feature selection techniques including SelectkBest, Recursive Feature Elimination, and Random Forest, this study concentrated on Fitbit data. Health Monitoring devices are essential for tracking

daily activities and health indicators. According to the study, the Random Forest approach produced the highest accuracy, up to 85%.

Virgin Pulse

Virgin Pulse is a well-being service for workers that businesses can use to enhance workplace regulations, encourage worker involvement, and encourage healthy lifestyle choices. It utilizes cloud computing. Digital therapeutic interventions, medical and condition management, health coaching, and well-being involvement are all significant components.

Administrators can collect employee data from a variety of sources, such as biometrics, health assessments, platform usage, habits, hobbies, and activities, using Virgin Pulse's real-time program analytics module. The data can then be analyzed to gain new insights. Employees receive a customized dashboard with annual care recommendations, customized reports with advice on preventing disease and health situations, regular reminders about healthy practices, and more.

Technology Used:

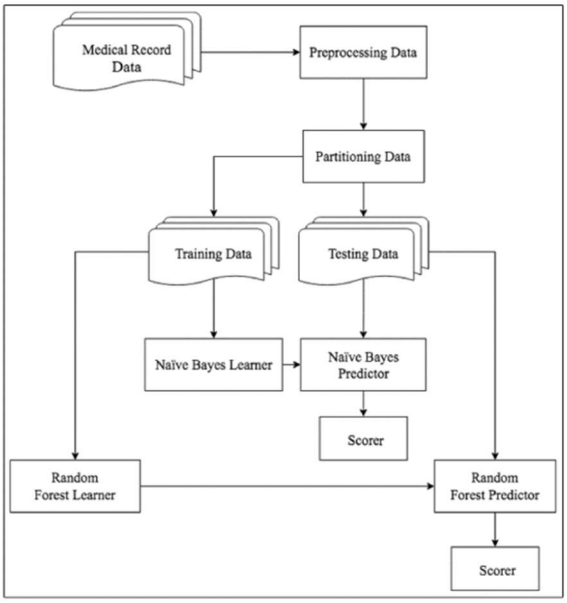
Java, PostgreSQL, AngularJS, Redis, Spring, Cassandra, Hibernate, Pusher, Liquibase, Filestack, Twilio SendGrid, sendwithus, npm, Bitbucket, Grunt, TeamCity, Jira

Text Mining in Healthcare for Disease Classification using Machine Learning Algorithm

Every day, enormous volumes of data are produced by cell phones and the development of technology, including vital healthcare data. Text processing is used in this study to analyze patient medical records from a public hospital from 2017 to 2019. A high accuracy

of almost 99.9% is achieved in disease categorization using the Random Forest and Naïve Bayes algorithms. The outcomes demonstrate the accuracy and robustness of the suggested strategy in categorizing medical record data with numerous disorders. From the massive volume of unstructured text data produced by information technology, text mining recovers knowledge. Electronic medical records are accumulating quickly in the healthcare industry, especially in public health services. This study uses machine learning to find the best categorization model through data gathering, processing, and analysis to improve disease prediction accuracy. Improving clinical management and diagnosis prediction within the framework of large medical record sets is the main goal.

Utilizing 2271 records spanning 19 disease categories, the study uses medical record data from General Hospital Dr. Soetomo Surabaya. The records span the period from February 1, 2017, to February 1, 2019. Steps in the methodology involve data collection, preprocessing, classification procedures, and evaluation.

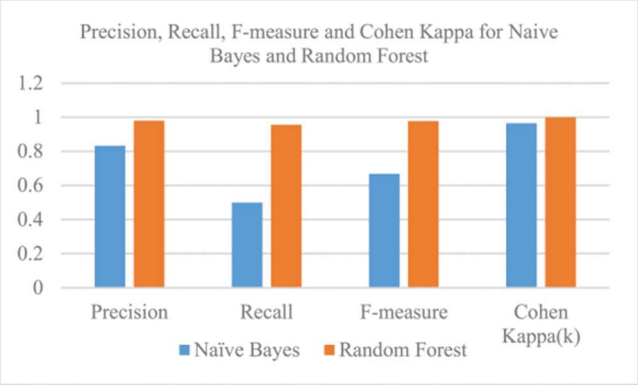


A Random Forest (RF) model that makes use of the Random Forest Learner and Predictor nodes is used in the illness categorization process. Several decision trees, each examined on various sets of rows and columns, are used to teach the RF Learner. The resulting model shows a Random Forest enforced by a majority vote at appropriate predictor nodes. The accuracy of disease classification is then improved by the RF Predictor node, which makes predictions based on the combined predictions of individual trees. The Scorer node and Confusion Matrix are used to evaluate metrics including True Positive Rate, False Positive Rate, Precision, Recall, F-Measure, Accuracy, and Cohen's kappa.

The study compares Naïve Bayes (NB) with Random Forest (RF) for disease categorization using KNIME machine learning. Using decision trees, RF performs better than NB, with an astounding accuracy of 99.9% as opposed to 97.2% for NB. The better precision, recall, F-measure, and Cohen's Kappa values of RF demonstrate its superior performance in disease classification using data from medical records. The confusion matrix exhibits the effectiveness of RF in both true positives and true negatives, underscoring its supremacy. All things considered, RF is the best model since it can accurately predict diseases in this health data mining study.

Classification Models	Correctly Classified	Incorrectly Classified
Naïve Bayes	97.2%	2.76%
Random Forest	99.9%	0.03%

Classification Models	Precision	Recall	F-measure	Cohen Kappa(k)
Naïve Bayes	0.931	0.523	0.687	0.968
Random Forest	0.979	0.955	0.977	1



5. Brief Thoughts on Similar Systems:

1. Covid-19 Data System:

Excellent application of Python crawler technology for real-time epidemic data; emphasizes data analytics in crucial healthcare scenarios, stressing the importance of data-driven decision-making in emergencies.

2. Heartbeat Monitoring with Machine Learning:

Predicting problematic FitBit patterns using machine learning (ML), such as Random Forest, highlights the promise of predictive analytics in personal health monitoring. The necessity of trustworthy algorithms in health applications is highlighted by the focus on accuracy rates.

3. Virgin Pulse Well-being Service:

Reflects the growing trend of incorporating technology into workplace wellness initiatives; real-time analytics and cloud computing are used to promote healthy lifestyle choices that are consistent with the preventative healthcare philosophy.

4. Text Mining for Disease Classification:

Outstanding accomplishment of almost 99.9% classification accuracy for diseases; a comparison of Random Forest and Naïve Bayes models shows why the latter performs better when handling big datasets of medical records.

6. Comparison of my application with Similar systems

Features/Criteria	My System	Big Data Crawling and Visualization System (COVID-19)	Machine Learning for Heartbeat Monitoring	Virgin Pulse Well-being Service	Text Mining in Healthcare for Disease Classification
Platform Type	Web App	Web App	Wearable Device	Cloud-Based Well-being Service	Research (Machine Learning Algorithm)
User Accessibility	Web Browsers	Web Browsers	Wearable Device App	Cloud Platform	Web Browsers
Predictive AI Models	Yes	Yes	Yes	Yes	Yes
Patient Engagement	Yes	Yes	Yes	Yes	No
AI for Health Trends and Risks Forecasting	Yes	Yes	No	Yes	No
Data-Driven Healthcare Decision-Making	Yes	No	Yes	Yes	No
User-Friendly Interface	Yes	Yes	No	Yes	No
Predictive Analysis Using AI Algorithms	Yes	Yes	Yes	Yes	No

Natural Language Processing	Yes	No	No	No	Yes
Data Retrieval Mechanisms	Yes	Yes	No	Yes	No
Robust Data Storage	Yes	Yes	No	Yes	No

7. Findings of research from similar systems

The goal of the healthcare data analytics platform is to provide medical practitioners with valuable insights by applying predictive modeling and trend analysis. The web-based tool supports data-driven decision-making for businesses and healthcare professionals by facilitating a smooth integration into hospital workflows. Predictive algorithms and real-time analytics are crucial, as demonstrated by similar systems like Heartbeat Monitoring using Machine Learning and the COVID-19 Data System. The platform is notable for its extensive feature set, which includes robust data storage, natural language processing, predictive analysis, and an intuitive UI. Its congruence with industry trends is highlighted by the comparative analysis, which also positions it as a useful tool for better patient outcomes and educated healthcare decision-making.

V. Planning

A. Methodology

In the subject of project management, there are various strategies that strive to boost productivity. To ensure project success, enhance teamwork, and make the most use of the resources at hand, effective project management strategies are crucial. The optimal framework might range from traditional approaches like Waterfall to modern agile alternatives, depending on the nature of the work, team dynamics, and project requirements. With the benefits that each technique provides, groups can tailor their strategy and finish tasks faster.

- Scrum

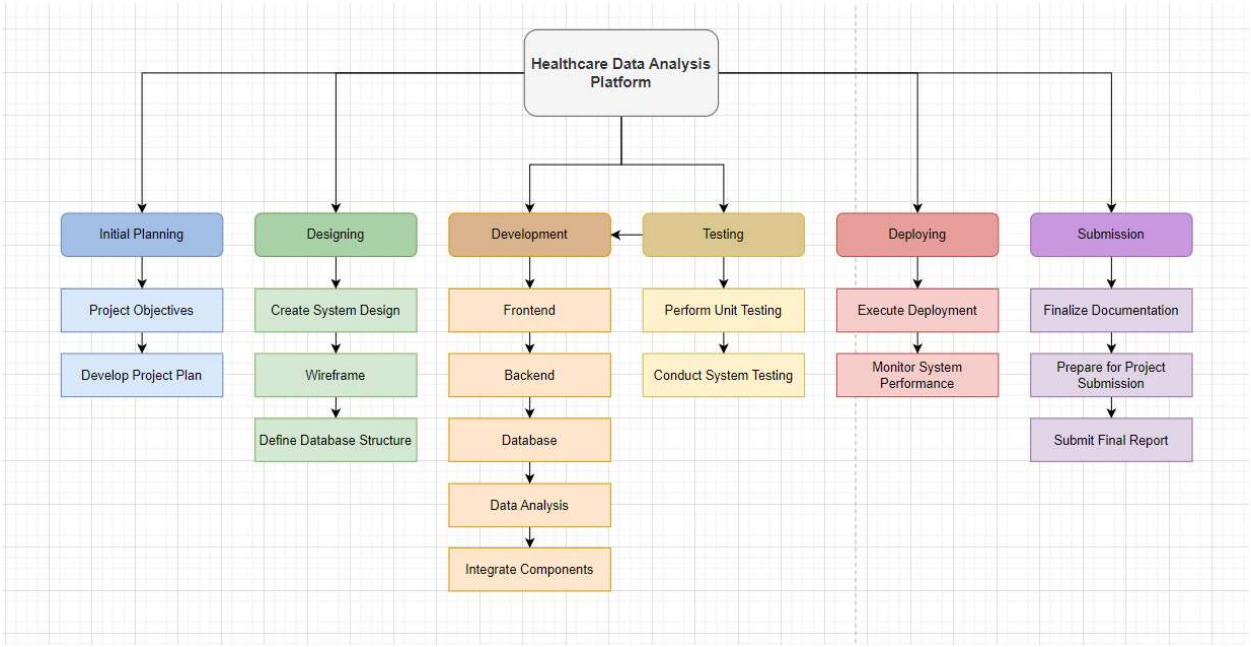
  Scrum is an improvement over Agile Management. A set of clearly defined roles and procedures that must be adhered to throughout the software development lifecycle

form the basis of the scrum methodology. Sprints are short, recurring work units used in Scrum that are also a synonym for feedback and reflection. These blocks usually consist of two or four weeks. Each sprint consists of a complete result, the best possible version of the final product that can be delivered to the customer with the least amount of effort upon request. Put otherwise, every Sprint exists as a unique entity.

B. Work Breakdown Structure (WBS)

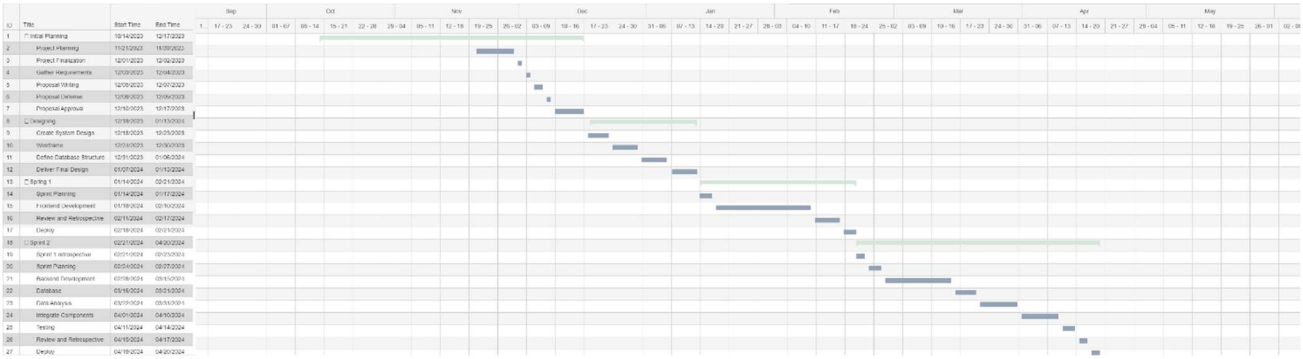
The work breakdown structure (WBS) is a project's classified split into manageable and precisely specified sub-parts. It establishes the overall outline of the project while also organizing it. It gives an ordered picture of the activity's scope by acting as a visual guide that divides the entire task into smaller more tractable ones. The fundamental purpose of designing this type of design (WBS) is to streamline large projects for better planning, implementation, and management of features that would otherwise be overwhelming. With adequate organization brought about by Work Breakdown Structure (WBS), knowledge regarding work allocation gets smoother, allowing for greater

regulation over procedures on management elements and so driving toward achievement in completing deadlines with efficiency.



B. Gantt Chart

The Gantt chart is an excellent tool for project management, substantially easing the creation and planning of any task magnitude. It shines notably when it comes to visualizing the growth of a project. Project managers can obtain insight into their progress by using this graphic tool that plots activities against time constraints. Gantt charts in handling software take a step further by improving the accuracy with which tasks are assigned, all while taking into account details such as team engagements and historical and present availability.



IV. Artifacts

A. Sub-process about Application (Web)

User Authentication and Authorization:

- Provide a safe login process where users, such as administrators, medical professionals, or ordinary users, can register and gain access to the platform according to their roles.

Data Input and Validation

- Create intuitive user interfaces for healthcare data entry, and put validation procedures in place to guarantee the data's integrity and accuracy.

Data Processing and Storage:

- Create server-side logic, such as validation and normalization, to handle incoming healthcare data. Put this information in a secure SQL database.

API Endpoints:

- Establish a series of clearly defined API endpoints to enable RESTful data retrieval, submission, and changes between the front end and the server.

Data Cleaning and Preprocessing:

- Include in place algorithms to deal with missing data, find and fix anomalies, and carry out feature engineering to improve the quality of the data that is utilized for analysis.

Machine Learning Integration:

- Selecting and putting into practice machine learning algorithms appropriate for predictive analytics will enable the system to identify trends in previously collected healthcare data.

Predictive Analytics:

- Create procedures that use machine learning models that have been taught to forecast outcomes based on fresh medical data, offering insightful information for making decisions.

. Reporting and Visualization:

- Create interactive dashboards that let users view and analyze medical data. For efficient data exploration, make sure there are real-time updates and responsiveness.

Security Measures:

- Use encryption to safeguard private medical information both during storage and transfer. Implement access controls to limit user access according to roles.

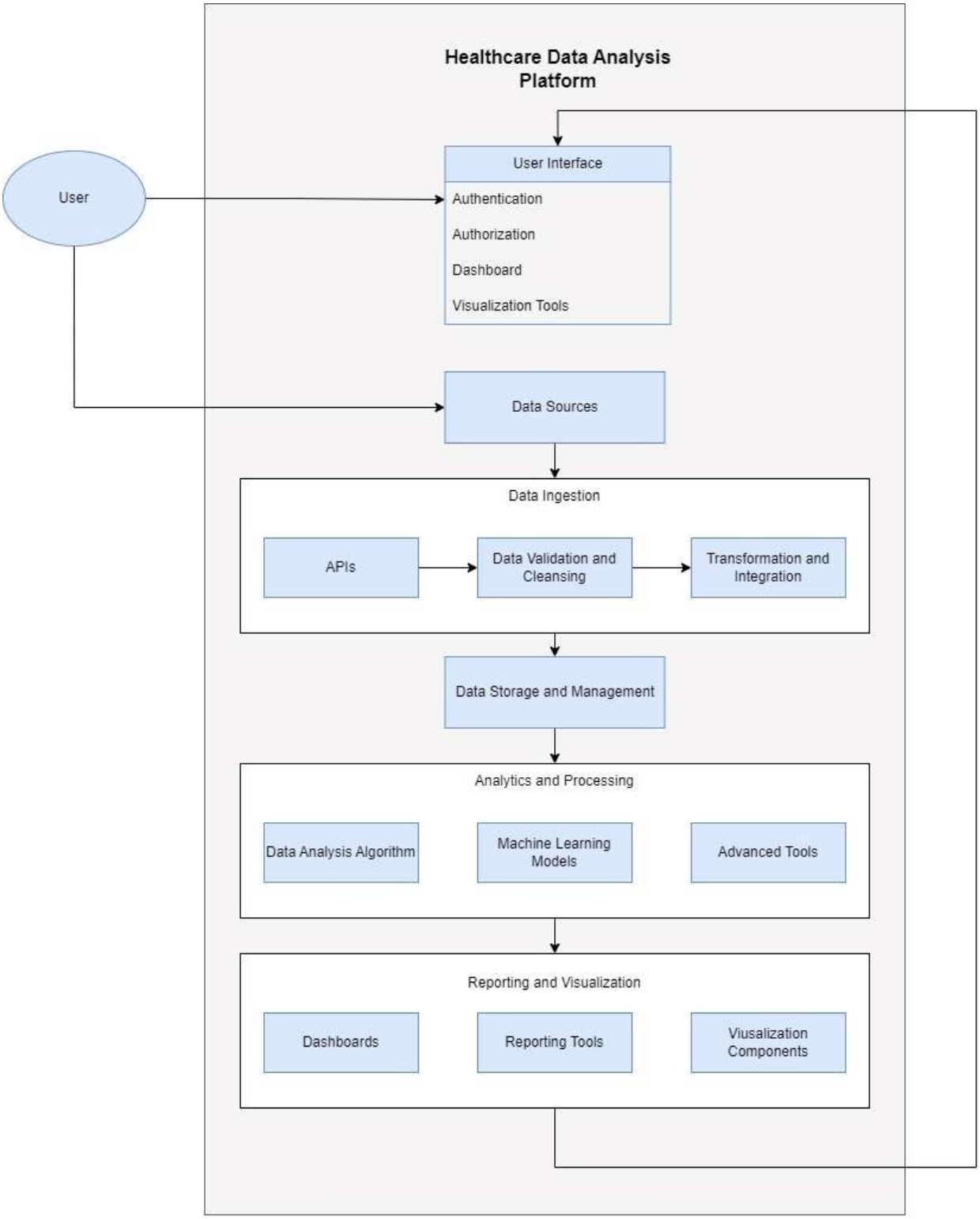
Notifications and Alerts:

- Use alerts and notifications to notify users of any significant platform changes, system updates, or critical occurrences.

Dashboard Interactivity:

- Add interactive components to dashboards, such as date pickers and filters, to let users explore and evaluate healthcare data in real time. Update data in real time to gain relevant insights.

C. Block Diagram of the System Workflow


Healthcare Data Analysis Platform
User
User Interface
Authentication
Authorization
Dashboard
Visualization Tools
Data Sources
Data Ingestion
APIs
Data Validation and Cleansing
Transformation and Integration
Data Storage and Management
Analytics and Processing
Data Analysis Algorithm
Machine Learning Models
Advanced Tools
Reporting and Visualization
Dashboards
Reporting Tools
Viusalization Components

SRS

This artifact report provides a comprehensive overview of the Software Requirements Specification (SRS) document created for the EnVIIsage Healthcare Data Analytics Platform project. The SRS captures critical system requirements to guide the development of this platform for healthcare data analytics and AI.

Purpose

The purpose of creating the SRS for the EnVIIsage Healthcare Data Analytics Platform is to clearly define the capabilities and requirements for the system to be developed. Having a complete and well-documented SRS ensures that all stakeholders - including the customer, development team, and users - have a shared understanding of what the system should do.

The SRS acts as the foundation and reference for all development activities. It guides the design, coding, testing, and release of the system. Defining requirements upfront through the SRS reduces ambiguity, aligns priorities, and sets expectations on deliverables to avoid issues down the line.

Scope

The scope of the EnVIIsage system covers the development of a web-based application for healthcare data analytics. Key capabilities in scope include:

- User management with authentication and authorization
- Healthcare data input, processing, and storage
- API endpoints for integration with other systems
- Data analytics through machine learning models
- Interactive reporting and dashboards
- Patient health predictions and risk analysis

Out of scope are capabilities like real-time data streaming, mobile app development, and connectivity with wearables/IOT devices. These can be considered in future releases.

Defining the scope enables focused effort on the must-have functionality for the first release and avoids wasting effort on nice-to-have features early on.

Overview of System Modules

The system consists of multiple modules, including but not limited to:

- User Authentication and Authorization
- Data Input and Validation
- Data Processing and Storage
- API Endpoints
- Machine Learning Integration
- Predictive Analytics
- Reporting and Visualization
- Security Measures
- Notifications and Alerts
- Dashboard Interactivity

Each module plays a crucial role in ensuring the overall functionality and effectiveness of the system.

Requirements Gathering Process

Requirements for the EnVIIsage platform were gathered through a combination of research papers, data analysis projects, literature reviews, and similar systems.

- Reviewed research papers from academic journals and conferences related to healthcare data analytics to understand the current state of technology and innovations in this space.
- Researched similar healthcare analytics solutions available to analyze their features and functionalities. This helped me identify gaps and areas of improvement for my project.
- Studied published case studies of healthcare analytics projects to gain insights into real-world challenges and lessons learned that are useful to consider.
- Had discussions with my project advisor to get their expert feedback on potential features and viability of my high-level solution ideas.
- Used user stories and use case modeling to capture the requirements from an end-user standpoint rather than my own assumptions as a developer.

- Created UI mockups and prototypes to visualize the workflows and get better feedback from users.

Prioritization Schema

The following prioritization schema was defined to categorize requirements based on their level of importance:

MUST (M) - Critical requirements without which the system would be unusable. Highest priority.

SHOULD (S) - Important requirements adding significant value but not absolutely essential. Medium priority.

COULD (C) - Nice-to-have requirements that improve user experience. Lowest priority.

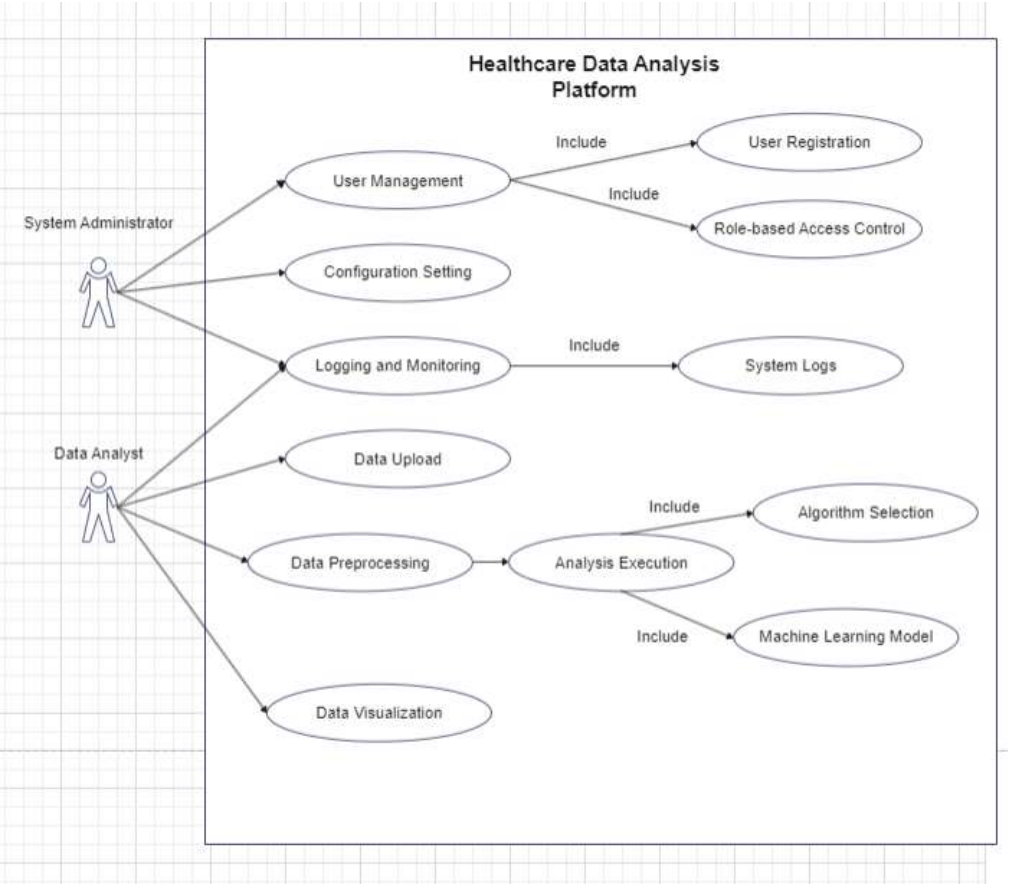
WON'T (W) - Out-of-scope requirements recorded for consideration in future releases.

This schema enabled structured thinking on the core MVP versus additional features.

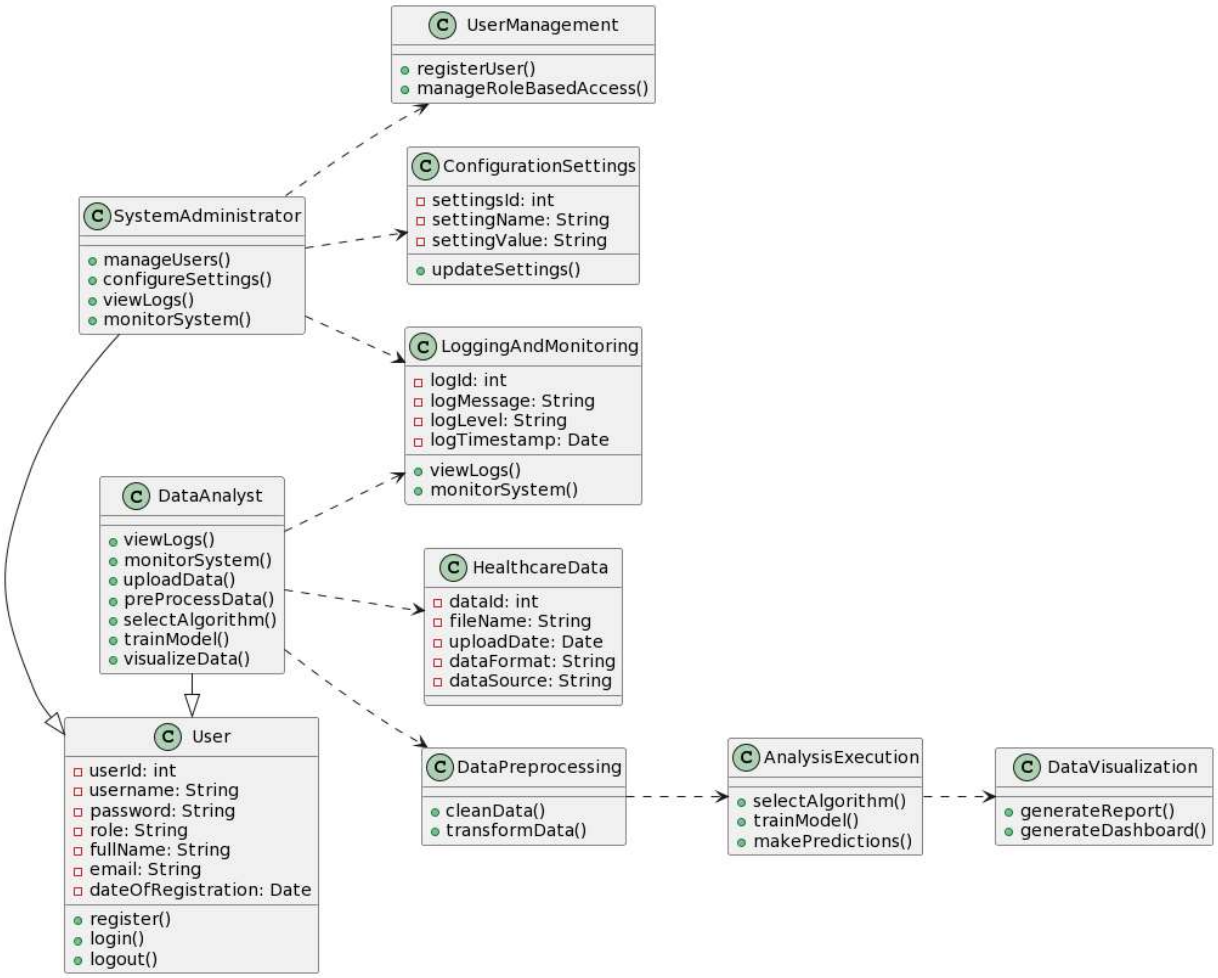
Requirement	Priority	Description
Functional Requirement		
Health Data Aggregation and Management	M	Aggregate and manage structured and unstructured health data from various sources, including demographics, medical history, diagnoses, lab tests, and imaging data, for organizational use. Validate and store data in a centralized repository.
Real-time Data Analytics and Visualization	M	Provide real-time analytics and visualizations for aggregate population cohorts, identifying public health trends, epidemiological mapping, and demographic analysis. Present through interactive visualizations and dashboards.
Predictive Risk Analysis and Alerts	S	Utilize machine learning models on aggregated data to predict public health risks such as disease outbreaks. Generate alerts for healthcare organizations and authorities.
Custom Reporting and Decision Support	C	Enable organizations to generate customized reports on public health trends and risk analysis to assist in decision-making and resource allocation. Provide recommendations based on analytics.
Non-Functional Requirement		

Requirement	Priority	Description
User Authentication and Access Control	M	Implement secure user authentication with role-based access control to health data based on organizational roles and permissions.
Data Privacy and Security	M	Ensure data privacy and security by encrypting data both at rest and in transit. Anonymize sensitive data used for analytics. Comply with relevant healthcare data regulations.
Scalable Architecture	S	Design the system to scale with increased data volumes and users. Leverage cloud infrastructure for scalability and flexibility.
High Availability and Reliability	W	Maintain high availability and reliability by implementing redundancy and failover mechanisms. Aim for minimal downtime and ensure data integrity.
Usability Requirement		
Intuitive and Responsive UI	M	Provide an intuitive and responsive user interface for organizations to interact with health data and analytics. Ensure ease of use and accessibility.
Contextual Help and Documentation	M	Include contextual help features and searchable documentation to assist users in navigating and utilizing the platform effectively.
Dark Mode	C	Offer a dark mode option for improved visibility and reduced eye strain, particularly in low-light environments commonly found in healthcare settings.

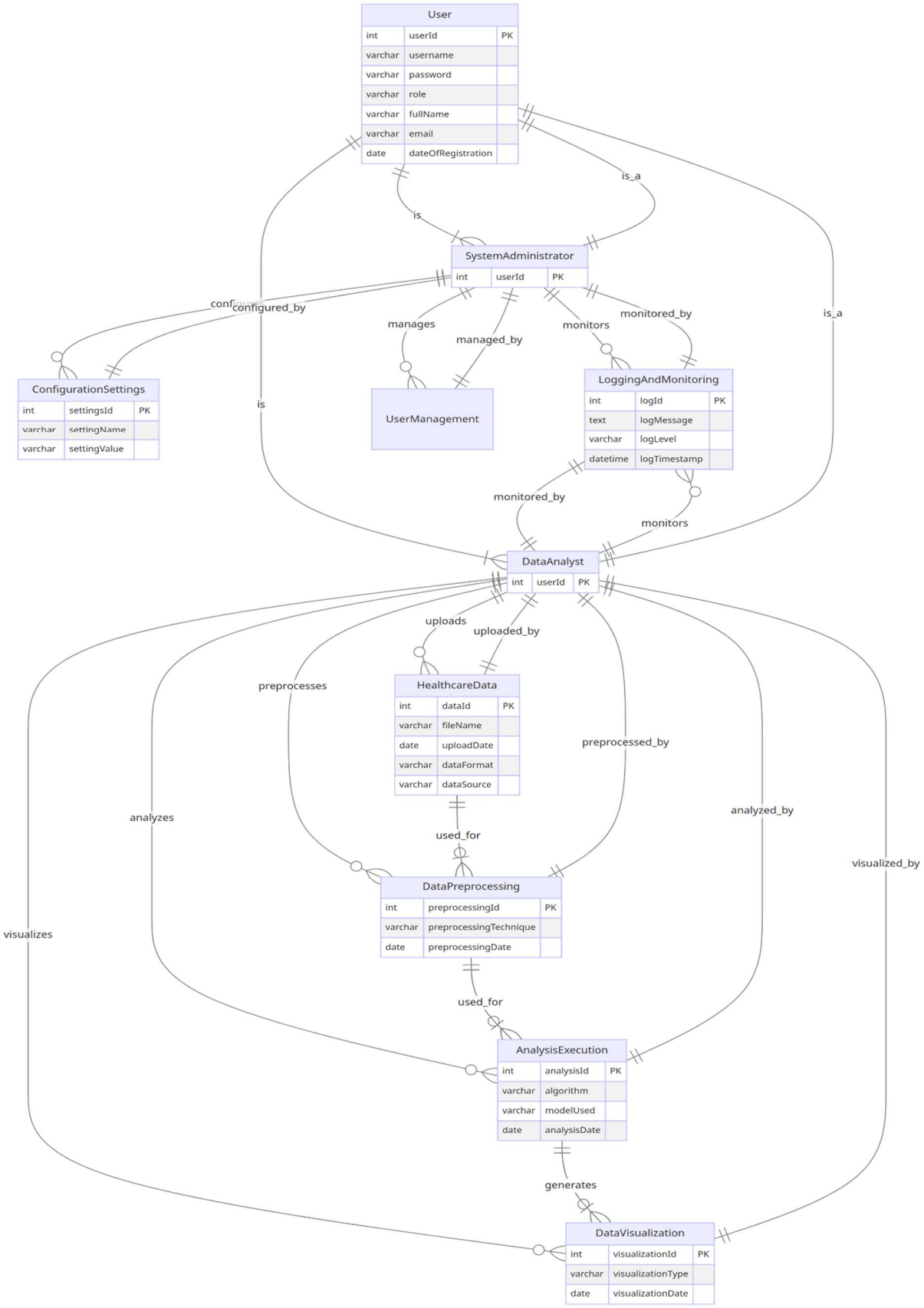
Use-case Diagram



Class Diagram


UserManagement
+ registerUser()
+ manageRoleBasedAccess()
ConfigurationSettings
- settingsId: int
- settingName: String
- settingValue: String
+ updateSettings()
SystemAdministrator
+ manageUsers()
+ configureSettings()
+ viewLogs()
+ monitorSystem()
LoggingAndMonitoring
- logId: int
- logMessage: String
- logLevel: String
- logTimestamp: Date
+ viewLogs()
+ monitorSystem()
DataAnalyst
+ viewLogs()
+ monitorSystem()
+ uploadData()
+ preProcessData()
+ selectAlgorithm()
+ trainModel()
+ visualizeData()
HealthcareData
- dataId: int
- fileName: String
- uploadDate: Date
- dataFormat: String
- dataSource: String
User
- userId: int
- username: String
- password: String
- role: String
- fullName: String
- email: String
- dateOfRegistration: Date
+ register()
+ login()
+ logout()
DataPreprocessing
+ cleanData()
+ transformData()
AnalysisExecution
+ selectAlgorithm()
+ trainModel()
+ makePredictions()
DataVisualization
+ generateReport()
+ generateDashboard()

Enitity-Relationship Diagram


User
int userId PK
varchar username
varchar password
varchar role
varchar fullName
varchar email
date dateOfRegistration
is_a
is
SystemAdministrator
int userId PK
configured_by
manages
managed_by
monitors
monitored_by
is_a
ConfigurationSettings
int settingsId PK
varchar settingName
varchar settingValue
is
UserManagement
LoggingAndMonitoring
int logId PK
text logMessage
varchar logLevel
datetime logTimestamp
monitored_by
monitors
DataAnalyst
int userId PK
uploads
uploaded_by
preprocesses
preprocessed_by
analyzes
analyzed_by
visualizes
visualized_by
HealthcareData
int dataId PK
varchar fileName
date uploadDate
varchar dataFormat
varchar dataSource
used_for
DataPreprocessing
int preprocessingId PK
varchar preprocessingTechnique
date preprocessingDate
used_for
AnalysisExecution
int analysisId PK
varchar algorithm
varchar modelUsed
date analysisDate
generates
DataVisualization
int visualizationId PK
varchar visualizationType
date visualizationDate

Data Dictionary

PK/FK	Field Name	Data Type	Size	Caption	Example
PK	userId	int	-	Unique user identifier	1, 2, 3
-	username	varchar	50	User's username	"user123"
-	password	varchar	50	User's password	"password123"
-	role	varchar	20	User's role	"Admin", "DataAnalyst"
-	fullName	varchar	100	User's full name	"Enviisage Platform"
-	email	varchar	100	User's email	"enviisage@gmail.com"
-	dateOfRegistration	date	-	User's registration date	"2023-05-01"
PK	settingsId	int	-	Unique setting identifier	1, 2, 3
-	settingName	varchar	50	Name of the setting	"MaxFileSize", "LogLevel"
-	settingValue	varchar	100	Value of the setting	"10MB", "INFO"
PK	logId	int	-	Unique log identifier	1, 2, 3
-	logMessage	text	-	Log message	"System started successfully"
-	logLevel	varchar	20	Log level	"INFO", "WARNING", "ERROR"
-	logTimestamp	datetime	-	Log timestamp	"2023-05-01 10:30:00"
PK	dataId	int	-	Unique data identifier	1, 2, 3

PK/FK	Field Name	Data Type	Size	Caption	Example
-	fileName	varchar	100	Name of the data file	"covid_data.csv"
-	uploadDate	date	-	Date of data upload	"2023-05-01"
-	dataFormat	varchar	20	Format of the data	"CSV", "Excel"
-	dataSource	varchar	100	Source of the data	"WHO", "CDC"
PK	analysisId	int	-	Unique analysis identifier	1, 2, 3
-	analysisType	varchar	50	Type of analysis	"Regression", "Classification"
-	modelUsed	varchar	50	Machine learning model used	"LinearRegression", "RandomForest"
-	analysisDate	date	-	Date of analysis execution	"2023-05-02"
PK	visualizationId	int	-	Unique visualization identifier	1, 2, 3
-	visualizationType	varchar	50	Type of visualization	"LineChart", "BarChart", "Report"
-	visualizationDate	date	-	Date of visualization generation	"2023-05-03"

In summary, this SRS provides a solid foundation to drive the development, testing, and release of the minimal viable product for the EnVIIsage Healthcare Data Analytics Platform in alignment with stakeholder needs. The requirements can evolve in future sprints based on user feedback.

VI. Conclusion

In conclusion, the Healthcare Data Analytics Platform is a ground-breaking strategy that aims to fundamentally transform the healthcare sector through the application of AI and data analytics. This project addresses significant scholarly challenges while providing patients and healthcare providers with a state-of-the-art, user-friendly platform. Individualized care and preventive health management are made possible by overcoming challenges in data analysis. The platform is committed to ongoing development as seen by its readiness to collaborate with healthcare institutions and own up to any errors. In essence, this project lays the foundation for modifications to healthcare practices that will enhance patient outcomes and facilitate better informed decision-making.

Several of the objectives of this research include the use of AI models, natural language processing, and dependable data storage systems. This comprehensive approach helps healthcare professionals to give patients individualized care while also assisting them in researching patterns, projecting results, and making informed decisions by spotting early warning indicators and possible health risks.

VII. References

A. Sundararaman, "Data Platform To Accelerate Healthcare Insights Generation," 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), Houston, TX, USA, 2023, pp. 470-475, doi: 10.1109/ICHI57859.2023.00069.

W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential", Health information science and systems, vol. 2, no. 1, pp. 3, 2014.

U. Akhtar, J. W. Lee, H. S. Muhammad Bilal, T. Ali, W. A. Khan and S. Lee, "The Impact of Big Data In Healthcare Analytics," 2020 International Conference on Information Networking (ICOIN), Barcelona, Spain, 2020, pp.

Project Management Institute. (2017). A Guide to the Project Management Body of Knowledge (PMBOK® Guide) - Sixth Edition. Newtown Square, PA: Project Management Institute.

GetApp, 2023. *GetApp.* [Online]
Available at: https://www.getapp.co.uk/software/106399/virgin-pulse#media
[Accessed 9 12 2023].

GetApp, 2023. *GetApp.* [Online]
Available at: https://www.getapp.co.uk/software/2066242/logi-symphony
[Accessed 10 12 2023].

nimblework, 2023. *nimblework.* [Online]
Available at: https://www.nimblework.com/agile/scruam-methodology/
[Accessed 10 December 2023].

Software Advice, 2023. *Software Advice.* [Online]
Available at: https://www.softwareadvice.com/corporate-wellness/virgin-pulse-profile/
[Accessed 3 12 2023].

Stackshare, 2023. *Stackshare.* [Online]
Available at: https://stackshare.io/virgin-pulse/virgin-pulse
[Accessed 3 December 2023].

Kumaresan, P. K. K. a. M., 2022. IEEE Xplore. [Online]
Available at: https://ieeexplore.ieee.org/document/9673220/references#references
[Accessed 3 January 2023].

U. Hayat, Z. I. A. M. a. S. A., 2023. IEEE Xplore. [Online]
Available at: https://ieeexplore.ieee.org/document/10139869/references#references

Kourou Konstantina et al., "Machine learning applications in cancer prognosis and prediction", Computational and structural biotechnology journal, vol. 13, pp. 8-17, Nov. 2014.

Nonso Nnamoko and Ioannis Korkontzelos, "Efficient treatment of outliers and class imbalance for diabetes prediction", Artificial Intelligence in Medicine, vol. 104, pp. 101-815, 2020, [online] Available: https://doi.org/10.1016/j.artmed.2020.101815.

Senerath Mudalige Don Alexis Chinthaka Jayatilake and Gamage Upeksha, "Involvement of Machine Learning Tools in Healthcare Decision Making", vol. 2021, [online] Available: https://doi.org/10.1155/2021/6679512.

N. N. Thilakarathne, W. D. M. Priyashan and C. P. Premarathna, "Artificial Intelligence - Enabled IoT for Health and Wellbeing Monitoring," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 01-07, doi: 10.1109/ICCCNT51525.2021.9579792. [online] Available: https://ieeexplore.ieee.org/document/9579792

A. Ed-daoudy and K. Maalmi, "A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment", Journal of Big Data, vol. 6, no. 1, 2019.

P. Kumar and U. Devi Gandhi, "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases", Computers & Electrical Engineering, vol. 65, pp. 222-235, 2018.

F. Al-Dhief et al., "A Survey of Voice Pathology Surveillance Systems Based on Internet of Things and Machine Learning Algorithms", IEEE Access, vol. 8, pp. 64514-64533, 2020.

F. B. Rejab, K. Nouira and A. Trabelsi, "Health monitoring systems using machine learning techniques" in Intelligent Systems for Science and Information, Springer, Cham, pp. 423-440, 2014.

K. G. Sheela and A. R. Varghese, "Machine Learning based Health Monitoring System", Materials Today: Proceedings, vol. 24, pp. 1788-1794, 2020.

A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. De Pristo, K. Chou, et al., "A guide to deep learning in healthcare", Nature medicine, vol. 25, no. 1, pp. 24-29, 2019.

M. E. Chowdhury, A. Khandakar, Y. Qiblawey, M. B. I. Reaz, M. T. Islam and F. Touati, "Machine Learning in Wearable Biomedical Systems", Sports Science and Human Health-Different Approaches. IntechOpen, 2020.

L. Li and G. Liu, "In-hospital Mortality Prediction for ICU Patients on Large Healthcare MIMIC Datasets Using Class Imbalance Learning", 2020 5th IEEE Int. Conf. Big Data Anal. ICBDA 2020, pp. 90-93, 2020.

M. Hao, H. Li, G. Xu, Z. Liu and Z. Chen, "Privacy-aware and Resource-saving Collaborative Learning for Healthcare in Cloud Computing", IEEE Int. Conf. Commun, vol. 2020-June, 2020.

Q. Liu, K. Liao and Z. Wei, "Automatic Acceptance Prediction for Answers in Online Healthcare Community", Proc. - 2018 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2018, pp. 1262-1265, 2019.

P. Jatunarapit, K. Piromsopa and C. Charoeanlap, "Development of Thai text-mining model for classifying ICD-10 TM", Proc. 8th Int. Conf. Electron. Comput. Artif. Intell. ECAI 2016, pp. 2-7, 2017.

I. A. Tache, M. Dragoicea, E. S. Apostol and C. O. Taurica, "Text mining of medical records", 2019 7th E-Health Bioeng. Conf. EHB 2019, pp. 2019-2022, 2019.

M. Wiley and J. F. Wiley, "Advanced R statistical programming and data models: Analysis machine learning and visualization", Apress Media LLC, 2019.

Performance Measures: Cohen's Kappa statistic - The Data Scientist, April 2021, [online] Available: https://thedatascientist.com/performance-measures-cohens-kappa-statistic/.