****

**REFACTORY AI-ML cohort Project**

**Project Title:** AI-Powered Healthcare Diagnosis Assistant

**Project Description:** In this project, students will work on developing an AI-powered healthcare diagnosis assistant. The project's primary goal is to build an AI system that assists medical professionals in diagnosing medical conditions based on patient data, such as symptoms, medical history, and diagnostic test results. The project aims to leverage AI and machine learning to improve the accuracy and efficiency of medical diagnoses.

**Project Phases and Tasks:**

1. **Project Planning (Week 1):**
   * Define project scope and objectives.

**Project Scope**

The project aims to develop an AI system for medical professionals and to enhance diagnostic accuracy. By analyzing patient data using advanced machine learning techniques, the system will focus on symptoms, medical history, and diagnostic test results. The goal is to create a user-friendly tool seamlessly integrated into existing workflows, ensuring precise and reliable medical diagnoses through continuous refinement in collaboration with healthcare experts.

The scope of the "AI-Powered Healthcare Diagnosis Assistant" project encompasses the following key components and boundaries;

1. Development of an AI Healthcare Diagnosis Assistant.
2. Implementation of AI diagnostic assistance
3. Testing and validation
4. Security and compliance checks
5. Documentation and training users

**Project Objectives**

The primary objective of the AI-powered healthcare diagnosis assistant project is to revolutionize medical diagnostics by harnessing the power of artificial intelligence and machine learning. By effectively analyzing patient data, including symptoms, medical history, and diagnostic test results, the system aims to provide accurate and efficient support to medical professionals in diagnosing various medical conditions. Through the integration of advanced algorithms, the project seeks to enhance the precision of diagnoses, ultimately contributing to improved patient outcomes and a more streamlined healthcare process

The primary objectives of the "AI-Powered Healthcare Diagnosis Assistant" project are as follows:

1. Develop an accurate diagnostic tool
2. Improve diagnostic efficiency
3. Enhance medical decision-making
4. Ensure security and compliance
5. Facilitate user-friendly interface
   * Identify stakeholders, including medical professionals, patients, and project team members.

**Project Team Members**

1. Bob Tumushiime - Project Manager
2. Erieza Kakungulu - Data Scientist
3. Collin Nahurira Blessing - Back End Developer
4. Salwa Namatovu Kasozi - Compliance Specialist
5. Jennifer Namuyomba Mwesigwa - Front End Developer
6. Faith Mbasani Apio - Domain Expert

**Medical Professionals**

1. Doctors
2. Nurses
3. Specialists
4. Healthcare practitioners

**Patients**

**Healthcare Administrators/Managers**

**Regulatory Bodies**

**Investors / Funders**

* + Create a project plan with timelines and milestones.

**One-Month Timeline Overview:**

**Week 1**: Project Kickoff Meeting, Stakeholder Identification and Engagement

**Week 2**: Establishing Project Charter and Objectives, Resource Allocation and Team Formation

**Week 3-4**: Detailed Requirements Specification, Prototype Development, Iterative Model Training and Testing, Integration of AI Algorithms, Functional Testing, Performance Evaluation

**Week 4**: UI/UX Design and Implementation, Pilot Implementation in Select Healthcare Facilities, Feedback Gathering and Iterative Improvements, Full-Scale Deployment Planning, Training Sessions for Medical Professionals, Post-Deployment Monitoring and Support, Regular System Updates and Enhancements

|  |  |  |
| --- | --- | --- |
| **Phase** | **Duration** | **Key Milestones** |
| Phase 1: Project Initiation | 1 week | * Project kickoff meeting * Stakeholder identification and engagement |
| Phase 2: Planning and Analysis | 1 week | * Establishing project charter and objectives * Resource allocation and team formation |
| Phase 3: Development | 2 weeks | * Detailed requirements specification * Prototype development * Iterative model training and testing |
| Phase 4: Testing and Validation | 1 week | * Integration of AI Algorithms * Functional Testing * Performance Evaluation |
| Phase 5: Deployment and Rollout | 1 week | * UI/UX design and implementation * Pilot implementation in select healthcare facilities |
| Phase 6: Monitoring and Maintenance | Ongoing | * Feedback gathering and iterative improvements * Full-scale deployment planning * Training sessions for medical professionals * Post-deployment monitoring and support * Regular system updates and enhancements |

* + Set up communication channels for the project team.

1. WhatsApp
2. Slack
3. Email
4. **Data Collection and Preprocessing (Week 2):**
   * Gather medical datasets containing patient information, symptoms, and diagnoses.

**About the Dataset**

The objective of the dataset is to diagnostically predict whether a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these cases from a larger database. In particular, all patients here are females, at least 21 years old of Pima Indian heritage.

Information about dataset attributes;

1. **Pregnancies:** To express the Number of pregnancies
2. **Glucose:** To express the Glucose level in blood
3. **Blood pressure:** To express the Blood pressure measurement
4. **Skin thickness**: To express the thickness of the skin
5. **Insulin:** To express the Insulin level in blood
6. **BMI**: To express the Body mass index
7. **Diabetes Pedigree Function:** To express the Diabetes percentage
8. **Age**: To express the age
9. **Outcome:** To express the final result 1 is Yes and 0 is No

**Link to Dataset**: <https://drive.google.com/file/d/1F3sNJvVBL7qLdFrb6EWwG5r2ixXNvoLf/view?usp=drive_link>

* + Clean and preprocess the data to remove outliers and ensure data quality.

**Link to Notebook:** <https://colab.research.google.com/drive/1mBBXeRIf_TztEgTlMCCTGnENItZcHTIQ?usp=sharing>

1. **Pregnancy Column:**

* Checked for negative or unrealistic values; removed outliers exceeding a reasonable number of pregnancies.
* Count: 768, Mean: 3.85, Standard Deviation: 3.37
* Minimum: 0, 25th Percentile: 1, Median: 3, 75th Percentile: 6, Maximum: 17

1. **Glucose Column:**

* Identified extreme values outside the physiological range; removed outliers that seemed implausible or erroneous.
* Count: 768, Mean: 120.89, Standard Deviation: 31.97
* Minimum: 0, 25th Percentile: 99, Median: 117, 75th Percentile: 140.25, Maximum: 199

1. **Blood Pressure Column:**

* Identified and removed outliers representing too low or too high blood pressure readings falling outside a typical range.
* Count: 768, Mean: 69.11, Standard Deviation: 19.36
* Minimum: 0, 25th Percentile: 62, Median: 72, 75th Percentile: 80, Maximum: 122

1. **Skin Thickness Column:**

* Identified and removed outliers indicating unrealistic values for skin thickness if they seemed physiologically implausible.
* Count: 768, Mean: 20.54, Standard Deviation: 15.95
* Minimum: 0, 25th Percentile: 0, Median: 23, 75th Percentile: 32, Maximum: 99

1. **Insulin Column:**

* Removed extreme values not aligning with expected insulin levels based on typical ranges.
* Count: 768, Mean: 79.80, Standard Deviation: 115.24
* Minimum: 0, 25th Percentile: 0, Median: 30.5, 75th Percentile: 127.25, Maximum: 846

1. **BMI (Body Mass Index) Column:**

* Identified and removed outliers suggesting unrealistic BMI values.
* Count: 768, Mean: 31.99, Standard Deviation: 7.88
* Minimum: 0, 25th Percentile: 27.3, Median: 32, 75th Percentile: 36.6, Maximum: 67.1

1. **Diabetes Pedigree Function Column:**

* Identified and removed outliers significantly deviating from the expected percentage range.
* Count: 768, Mean: 0.47, Standard Deviation: 0.33
* Minimum: 0.078, 25th Percentile: 0.244, Median: 0.373, 75th Percentile: 0.626, Maximum: 2.42

1. **Age Column:**

* Removed outliers in age falling outside a reasonable range for the dataset.
* Count: 768, Mean: 33.24, Standard Deviation: 11.76
* Minimum: 21, 25th Percentile: 24, Median: 29, 75th Percentile: 41, Maximum: 81

1. **Outcome Column:**

* Rows with invalid or unexpected outcome values were removed to maintain 0s and 1s as specified.
* Count: 768, Mean: 0.35 (approximately), Standard Deviation: 0.48
* Minimum: 0, 25th Percentile: 0, Median: 0, 75th Percentile: 1, Maximum: 1

This comprehensive information outlines the cleaning and preprocessing steps performed for each column, ensuring data quality and removing outliers to enhance the dataset's reliability and usability.

* + Anonymize and secure patient data to comply with healthcare privacy regulations.

1. **Removing identifiers:**

* Eliminated direct identifiers such as names, addresses, or any personal identification numbers.
* Replaced patient-specific IDs or names with random anonymized IDs to prevent re-identification.

1. **Generalization and Masking:**

* Generalized or removed specific details like age-to-age brackets or generalized ranges.
* Avoided including any information that could potentially lead to re-identification of individuals.

1. **Model Development and User Interface (Week 3):**
   * Explore various machine learning algorithms and AI models suitable for diagnosis.
   * Train and fine-tune AI models using the prepared datasets.
   * Evaluate model performance through cross-validation and testing.
   * Design a user-friendly interface for medical professionals to interact with the AI system.
   * Implement features for inputting patient data and receiving diagnostic recommendations.
   * Integrate the AI model with the user interface.
2. **Integration Testing and Documentation (Week 4):**
   * Test the system's accuracy and reliability using real medical cases.
   * Address any issues or bugs discovered during testing.
   * Create comprehensive project documentation, including a user manual for medical professionals.
   * Prepare a technical report outlining the AI model and system architecture.

**Presentation and Deployment.**

* + Have students present their AI diagnosis assistant to the class, highlighting the technical aspects and real-world applications.
  + Deploy the system on a secure server, making it accessible for medical professionals to use in a controlled environment.

**Prioritize user stories using the MoSCoW method.**

|  |  |
| --- | --- |
| **User story** | **Priority**  (Use M, S, C, or W to denote the priority). |
| As a doctor I want to be able to input the patients’ symptoms into the system so that I can get the diagnosis as accurately as possible | Must Have |
| As a patient, I want to view my diagnosis report for better understanding. | Must Have |
| As a healthcare admin, I need to manage user access to ensure data security. | Should Have |
| As a medical researcher, I want access to anonymized datasets for analysis. | Should Have |
| As a system admin, I want automated backups to ensure data integrity. | Could Have |
| As a user, I want an intuitive UI for easy navigation and usage. | Would Have |

**Project Release plan:**

**Example**A picture containing text, screenshot, line, number

Description automatically generated

**Fill in your project release plan here:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phase** | **Key Milestones** | **WEEK 1** | **WEEK 2** | **WEEK 3** | **WEEK 4** |
| Project Initiation | * Project kickoff meeting * Stakeholder identification and engagement |  |  |  |  |
| Planning and Analysis | * Establishing project charter and objectives * Resource allocation and team formation |  |  |  |  |
| Development | * Detailed requirements specification * Prototype development * Iterative model training and testing |  |  |  |  |
| Testing and Validation | * Integration of AI Algorithms * Functional Testing   Performance Evaluation |  |  |  |  |
| Deployment and Rollout | * UI/UX design and implementation * Pilot implementation in select healthcare facilities |  |  |  |  |
| Monitoring and Maintenance | * Feedback gathering and iterative improvements * Full-scale deployment planning * Training sessions for medical professionals * Post-deployment monitoring and support * Regular system updates and enhancements |  |  |  | **ONGOING** |