College of Engineering and Technology

Project Work – Student Hand Book

Project Batch ID : CTB07624

Degree/ program	B.Tech	**Specialisation**	Computer Science & Engineering
Academic Year	2024-2025 (Even)	**Semester**	8

Name of student	**Register Number**	**Department**	**Mobile Number**	**Email ID**
A.SIVA NARAYANA	RA2111003011444	C-TECH	9381444576	aa5710@srmist.edu.in
O.SAI RISHITHA	RA2111003011566	C-TECH	9666943340	oo2707@srmist.edu.in

Working Title of the Project:	METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM
Project Site / Location	Chennai
Name and address of the company / organisation (Applicable for projects with industry or industry support)	SRM IST, Kattankulathur, Chengalpattu District-603203

Supervision Team

	Supervisor	**Co-Supervisor**	**External Supervisor (If applicable)**
Name	Dr. S. SANKARA NARAYANAN		
Designation	**Assistant Professor**		
Department	**C-TECH**		
Campus	Kattankulathur		
Telephone	9976028293		
E-mail	sankaras1@srmist.edu.in		

Course Code	18CSP109L	Course Title	Metabolic syndrome detection and healthcare using genetic algorithm

Mission Statement

Problem (or) Product Description:

The "Metabolic Syndrome Prediction and Personalized Recommendation System" aims to develop an intelligent, data-driven solution for the early diagnosis and management of Metabolic Syndrome (MetS). This system will:

- Identify key risk factors: Employing advanced feature selection techniques to pinpoint the most relevant indicators of MetS from patient data.
- Predict MetS onset: Utilizing machine learning models to accurately predict the likelihood of an individual developing MetS.
- Assess MetS severity: Calculate a Metabolic Syndrome Severity Score to provide a nuanced understanding of the patient's condition beyond a binary diagnosis.
- Provide personalized recommendations: Generate patient-specific healthcare and lifestyle advice, including potential medication considerations based on severity classifications (high, medium, low).

The ultimate goal is to move beyond generalized approaches by offering a personalized solution for early MetS diagnosis and care, empowering individuals and healthcare providers with actionable insights.

Assumptions and Constraints

Assumptions:

- Data Availability: We assume access to a relevant dataset containing patient information related to the risk factors of MetS (e.g., blood pressure, blood sugar, waist circumference, cholesterol, triglycerides) and potentially demographic and lifestyle data. For Sprint 1, we might start with a publicly available dataset or a simulated dataset if real data access is pending.
- Initial Feature Set: We assume a preliminary set of features relevant to MetS can be identified based on medical knowledge and the information presented in the introduction.

- Focus on Core Functionality: Sprint 1 will primarily focus on the foundational steps of data exploration, preprocessing, and initial feature selection. Prediction model development and recommendation generation will be addressed in later sprints.
- Team Collaboration: We assume effective communication and collaboration among the team members.
- Basic Infrastructure: We assume the availability of necessary computing resources and software tools for data analysis and initial model development.

Constraints:

- Time Limit: Sprint 1 will be completed within a defined timeframe (e.g., 1-2 weeks).
- Limited Scope: The scope of Sprint 1 is limited to data understanding, preprocessing, and exploring initial feature selection techniques.
- Tooling: The team might be constrained by the available software libraries and tools.
- Initial Data Quality: The initial dataset might contain missing values, outliers, or inconsistencies that need to be addressed

Stakeholders

- Patients: Individuals who will benefit from early diagnosis, personalized recommendations, and improved management of MetS.
- Healthcare Providers (Doctors, Physicians): Professionals who will use the system to aid in diagnosis, treatment planning, and patient counseling.
- Researchers: Individuals interested in the development and validation of advanced machine learning techniques for metabolic health.
- Project Team: The individuals responsible for the development and implementation of the system.
- Potential Funding Agencies/Organizations: Entities that may support the project's development and deployment.

Division of work and contributors of SPRINT 1 [Include Daily Scrum of Sprint 1]

Overall Goal of Sprint 1: To understand the available data, perform initial preprocessing steps, and explore potential feature selection techniques.

Team Members:

- Siva Narayana: Data Acquisition, Exploration, and Initial Preprocessing
- Sai Rishitha: Feature Selection Technique Research and Initial Implementation

Division of Work:

- **Siva Narayana:**
    - Identify and acquire the initial dataset (publicly available or simulated).
    - Perform initial exploratory data analysis (EDA) to understand data characteristics, identify potential issues (missing values, data types), and gain insights into the variables.
    - Implement basic data cleaning procedures (handling obvious missing values, ensuring data consistency).
    - Document initial data findings and preprocessing steps.
- **Sai Rishitha:**
    - Research and understand different filter-based feature selection methods (e.g., correlation analysis, chi-squared test, mutual information).
    - Implement one or two basic filter methods on the initially cleaned data.
    - Evaluate the initial ranking of features obtained from these methods.
    - Document the findings and potential limitations of the explored methods.

Signature of the Supervisor

Division of work and contributors of SPRINT 2 [Include Daily Scrum of SPRINT 2]

Overall Goal of Sprint 2: To explore and implement more advanced feature selection techniques, including wrapper methods, and potentially begin initial model selection considerations.

- **Siva Narayana:** Advanced Data Preprocessing, Feature Engineering, and Feature Selection Integration
- **Sai Rishitha:** Wrapper Method Implementation and Evaluation
- **Siva Narayana:**
    - Based on the insights from Sprint 1, perform more advanced data preprocessing steps (e.g., feature scaling, handling skewed distributions).
    - Explore potential feature engineering opportunities based on domain knowledge and the literature review.
    - Integrate the preprocessed and potentially engineered data with Sai Rishitha's wrapper method experiments.
    - Oversee the evaluation of different feature subsets generated by the wrapper methods.
    - Document the overall feature selection process and the rationale behind chosen methods.
- **Sai Rishitha:**
    - Research and understand wrapper-based feature selection methods like Recursive Feature Elimination (RFE) and potentially begin exploring Genetic Algorithms (GA) conceptually.
    - Implement RFE using a basic machine learning model (e.g., Logistic Regression or a simple Tree-based model).
    - Evaluate the performance of the model with different subsets of features selected by RFE.

Signature of the Supervisor

Division of work and contributors of SPRINT 3 [Include Daily Scrum of Sprint 3]

Overall Goal of Sprint 3: To explore and implement the Genetic Algorithm (GA) for feature selection and begin initial prediction model exploration.

- **Siva Narayana**: Data Preparation, Model Input Structuring, and Initial Prediction Model Exploration
- **Sai Rishitha:** Genetic Algorithm (GA) Implementation and EvaluationDivision of Work:
- **Siva Narayana:**
    - Ensure the preprocessed and feature-engineered data is in the correct format for input into the prediction models.
    - Work on creating the structure for the Model Input Worksheet (as requested).
    - Prepare sample data instances for testing the prediction models.
    - Begin exploring the implementation of basic prediction models (e.g., Logistic Regression, Decision Trees) using the feature subsets identified in previous sprints.
- **Sai Rishitha:**
    - Implement the Genetic Algorithm for feature selection, defining the population, fitness function (using a chosen prediction model's performance), crossover, and mutation operators.
    - Run the GA to identify optimal or near-optimal subsets of features.
    - Evaluate the performance of a chosen prediction model using the feature subsets selected by the GA.
    - Document the GA implementation and results**.**

Signature of the Supervisor

Worksheet / Data collection / Observation etc

During Sprint 3, the focus on "Worksheet / Data collection / Observation" would involve:

- **Siva Narayana's efforts on the Model Input Worksheet:** This is a form of structured data collection for how model input will be organized. Observations here would be about the clarity and completeness of the worksheet design and the ease of populating it with sample data.
- **Sai Rishitha's GA experiments:** Observations would be recorded on:
    - How different GA parameters affect the feature selection process (e.g., convergence speed, diversity of feature subsets).
    - The performance of the chosen prediction model with the feature subsets selected by the GA.
    - The computational cost of running the GA with different configurations.
    - Any patterns or insights gained about which features the GA consistently selects as important.
- **General Data Collection/Observation:** The team would continue to observe the characteristics of the preprocessed data and how different feature selection methods impact the subsequent prediction model performance. Any unexpected data patterns or model behaviors would be documented.

Model Input Worksheet Example (Sample Format):

Patient ID	Age (years)	Systolic BP (mmHg)	Diastolic BP (mmHg)	Fasting Glucose (mg/dL)	Waist Circumference (cm)	HDL Cholesterol (mg/dL)
123	45	130	85	105	95	40
456	62	145	92	120	102	35
789	38	120	80	90	88	50

Research Article with Journal Publication Details / Patent disclosure form with patent status

Research Paper Successfully Submitted for Review - IJSREM Journal External Inbox ×

IJSREM Journal <editor@ijsrem.com> Mon, May 5, 1:37 PM (7 days ago)
to me

Dear ADDANKI SHIVA NNARAYANNA , O SAI RISHITHA , DR . S . SANKARA NARAYANAN,

Thanks for Submitting your Research Paper titled METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM. We will Review and update the status within 12-24 Hours. For more information feel free to reach us through email editor@ijsrem.com

1. **Manuscript Title**
   METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM
2. **Author Name's**
   ADDANKI SHIVA NNARAYANNA , O SAI RISHITHA , DR . S . SANKARA NARAYANAN
3. **Email Address**
   aa5710@srmist.edu.in
4. **Phone Number**
   9381444576
5. **Total No. of Authors**
   3
6. **Article Type**
   Research Paper