Thyroid Disease Detection

Group-F

Loyalist College in Victoria Park, Toronto Campus

2024S-T1 AISC1006 - Step Presentation (Step 1) 01 (M07 Group 1)

Prof. Usman Ahmad

May 25, 2024

Group Members

|  |  |
| --- | --- |
| Student Name | Student Id |
| Moksh Jaiswal | 500240046 |
| Alen Charuvila Saji | 500237019 |
| Adarsh Shriram Pednekar | 500233484 |
| Utsav Harshadbhai Khamar | 500238367 |
| Pranay Sai Jangeti | 500240045 |
| Taranjot Singh Bindra | 500239542 |
| Smit Rajendraprasad Patel | 500238279 |
| Om Kiranbhai Patel | 500228172 |
| Tanzima Mohammadyasin Shaikh | 500238212 |
| Aravind Seenivasan | 500236355 |

Abstract

The Thyroid Disease Detection project aims to enhance the accuracy and efficiency of thyroid disorder diagnosis using machine learning techniques. Leveraging advanced algorithms and data analytics, the project focuses on developing a robust and reliable system capable of analyzing diverse medical data, including thyroid function tests and imaging results. This system is intended to assist healthcare professionals in the early and accurate identification of thyroid disorders, leading to timely interventions and improved patient outcomes.

Project Goals

The primary objective of this project is to develop a machine learning-based solution for detecting thyroid disorders. The specific goals include:

* Enhancing the accuracy of thyroid disorder detection.
* Improving the efficiency of the diagnostic process.
* Assisting healthcare professionals in early and precise diagnosis.
* Facilitating timely interventions and improved patient care.

Assigned Roles

|  |  |
| --- | --- |
| Student Name | Roles |
| Moksh Jaiswal | Data Wrangling, EDA |
| Alen Charuvila Saji | Data Preprocessing, Reporting & PPT |
| Adarsh Shriram Pednekar | Model Deployment |
| Utsav Harshadbhai Khamar | Model Deployment |
| Pranay Sai Jangeti | Reporting and Documentation |
| Taranjot Singh Bindra | Introduction and Model Building |
| Smit Rajendraprasad Patel | Data Wrangling, EDA |
| Om Kiranbhai Patel | Conclusions and Findings |
| Tanzima Mohammadyasin Shaikh | Model Deployment |
| Aravind Seenivasan | Model Building |

Overall Tasks

To achieve the project objectives, the following tasks will be undertaken:

1. Data Collection and Preprocessing:
   1. Collecting the Thyroid Detection dataset from the UCI Machine Learning Repository and Kaggle.
   2. Cleaning and preprocessing the dataset to handle missing values, normalize data, and encode categorical variables.
2. Exploratory Data Analysis (EDA):
   1. Analyzing the dataset to understand the distribution and relationships between different attributes.
   2. Visualizing the data to identify patterns and correlations.
3. Feature Engineering:
   1. Selecting relevant features for model development.
   2. Creating new features if necessary to improve model performance.
4. Model Development:
   1. Choosing appropriate machine learning algorithms for classification.
   2. Training models using the preprocessed dataset.
   3. Evaluating model performance using metrics such as accuracy, precision, recall, and F1-score.
5. Model Validation and Testing:
   1. Validating the models using cross-validation techniques.
   2. Testing the models on a separate test set to assess their generalization capability.
6. Deployment:
   1. Implementing the best-performing model into a user-friendly system.
   2. Providing healthcare professionals with tools to input patient data and receive diagnostic predictions.
7. Documentation and Reporting:
   1. Documenting the entire process, including data preprocessing steps, model development, and evaluation results.
   2. Preparing a comprehensive report detailing the findings and outcomes of the project.

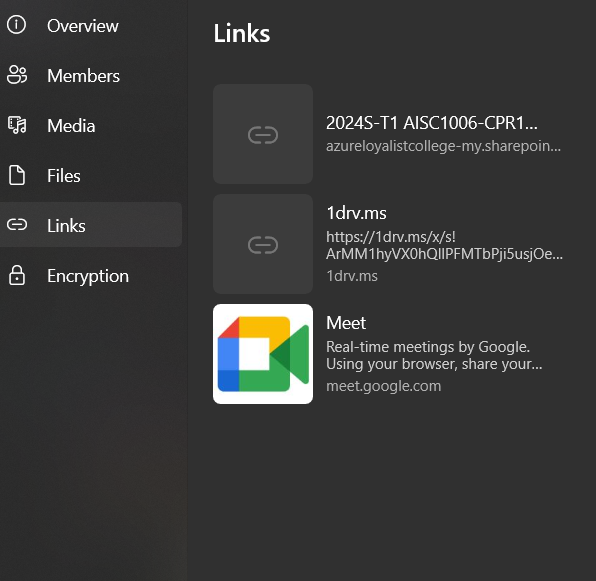
Project Deliverables

The project will result in the following deliverables:

* Preprocessed Dataset: A cleaned and prepared dataset ready for model training and testing.
* Exploratory Data Analysis Report: A report detailing the insights gained from data analysis.
* Machine Learning Models: Trained models capable of accurately detecting thyroid disorders.
* Model Evaluation Report: A detailed evaluation of model performance with metrics and visualizations.
* Deployed System: A user-friendly application or interface for healthcare professionals to use the developed models.
* Project Documentation: Comprehensive documentation covering all aspects of the project, from data preprocessing to model deployment.

**Proof of Discussion**

* We collectively participated in Google meet meetings, brainstormed ideas regarding the project and gathered the resources.



* After school, we met up to talk about how things were going. We followed a plan and split up the work. Right now, we finished cleaning up the data. Next, we're going to pick a model to train.

A screenshot of a computer

Description automatically generated

* In this meeting, we discussed the work done so far, and Moksh explained further developments regarding the project. We volunteered to take up tasks based on our interests for the project’s next steps, i.e., data preprocessing, model building, training the model, and model deployment.

A screenshot of a computer

Description automatically generated

**Model Building:**

**First Step:**

* Encoding: Convert Age and Boolean values to numerical format.
* One-hot encoding: For example, M = 0, F = 1. There are other encoding methods as well, such as label encoding.
* Data type conversion: Convert Boolean and categorical variables (e.g., sex) into numerical values.

**Second Step:**

* Feature Selection: Drop the unnecessary features that do not significantly impact the target variable.

**Third Step:**

* Feature Scaling: Normalize all attribute values to bring them into the same range.
* This ensures each feature (attribute) contributes equally to the model, preventing dominance by features with larger values.
* Common methods used are standardization and normalization.

**Step Four: Train-test split**

* Separate the dependent variable (target) and the independent variables.
* Create a train-test split.

**Project Management**

* **Repository:** <https://github.com/pranaysaaij/Thyroid-Disease-Detection-Group-F>
* **Project:** <https://github.com/users/pranaysaaij/projects/1>