# **Lab Cycle 2 - Mini Project**

**Natural Language Processing Laboratory(22MDCEL2)**

|  |  |  |
| --- | --- | --- |
| **S.No** | **Student Name** | **Roll Number** |
| **1)** | **Gokulanand M** | **71762233015** |
| **2)** | **SriKrishnan G** | **71762233049** |

**Abstract:**

The project **Marunthagam** aims to develop an intelligent chatbot that assists in prescribing medicines based on the user’s disease severity and the biological components of medications. The application will process informal queries from users, identify the symptoms or disease, and recommend suitable medicines by cross-referencing a knowledge base created from an optimized text corpus in the medical domain.

The knowledge base will be built from a comprehensive dataset of medical prescriptions and biological data of medicines. **Marunthagam** will also consult a virtual doctor component that reviews and confirms the prescribed medication. The chatbot will only display the medicine recommendation to the user if both the AI and the virtual doctor's suggestions align. In case of discrepancies, the best-suited prescription will be chosen based on efficacy.

**Major Concepts:**

* **Text Classification:** A classification algorithm such as Naive Bayes, Support Vector Machines (SVM), or Logistic Regression will categorize diseases based on input symptoms and disease severity. Additionally, a Voting Classifier can be employed, combining the strengths of multiple algorithms (e.g., Naive Bayes, SVM, Logistic Regression) to enhance accuracy in predicting the best medication.
* **Syntax Analysis:** Informal user queries will undergo syntactic parsing to analyze the structure and extract relevant medical terms.
* **Semantic Analysis:** Tools such as SpaCy or NLTK will be used to extract the semantic meaning of user input, ensuring the proper context and meaning are derived from the symptoms mentioned.

**Additional Features:**

* **Tagging Concepts:** Named Entity Recognition (NER) for tagging medical terms (diseases, symptoms, medications) and Part-of-Speech (POS) tagging will be applied to the user queries to improve information extraction.
* **Chunking and Information Extraction:** The system will extract chunks of information (symptoms, biological components of medicines) and identify relationships between disease severity and medication to make informed recommendations.