**Project Charter Document**

**Project Name:** Feature Extraction from Medical Journals.

**Department:** Health Care

**Focus Area:** Market Analysis

**Product/Process:** Data Analysis

**Project Team:** DS\_Project\_Team\_59



**Prepared By**

|  |  |
| --- | --- |
| **Document Owner(s)** | **Project/Organization Role** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Project Charter Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Change Description** |
|  |  |  |  |
|  |  |  |  |

**TABLE OF CONTENTS**

[**1**](#_heading=h.3whwml4) **PROJECT CHARTER PURPOSE 3**

[**2**](#_heading=h.2bn6wsx) **PROJECT EXECUTIVE SUMMARY 3**

[**3**](#_heading=h.qsh70q) **PROJECT OVERVIEW 3**

[**4**](#_heading=h.3as4poj) **PROJECT SCOPE 4**

[4.1](#_heading=h.1pxezwc) Goals and Objectives 4

[4.2](#_heading=h.49x2ik5) Project Deliverables 4

[4.3](#_heading=h.2p2csry) Deliverables Out of Scope 4

[4.4](#_heading=h.147n2zr) Project Duration 4

[**5**](#_heading=h.3o7alnk) **PROJECT CONDITIONS 5**

[5.1](#_heading=h.23ckvvd) Project Assumptions 5

[5.2](#_heading=h.ihv636) Project Issues 5

[5.3](#_heading=h.32hioqz) Project Risks 5

[5.4](#_heading=h.1hmsyys) Project Constraints 5

[**6**](#_heading=h.41mghml) **Project Structure Approach 6**

[**7**](#_heading=h.2grqrue) **Project Team Organization Plans 6**

[**8**](#_heading=h.vx1227) **PROJECT REFERENCES 6**

[**9**](#_heading=h.3fwokq0) **APPROVALS 6**

[**10**](#_heading=h.1v1yuxt) **APPENDICES 8**

[10.1](#_heading=h.4f1mdlm) Document Guidelines 8

[10.2](#_heading=h.2u6wntf) Project Charter Document Sections Omitted 8

# PROJECT CHARTER PURPOSE

* Manual annotation and feature extraction by medical experts are inefficient because they involve a complex process and require expensive and time-consuming labor. Therefore, efficient and accurate named entity recognition (NER) techniques are becoming increasingly important for use in computational data analysis to automatically analyze the biomedical literature and extract useful information from texts.



# PROJECT EXECUTIVE SUMMARY

* Project goals
* Objectives
* Scope
* Assumptions
* Risks
* Costs
* Timeline
* Approach
* Organization



# PROJECT OVERVIEW

* The increasing number of medical articles and resources, searching for and extracting valuable information has become challenging.
* Researchers consider multiple information sources and transform unstructured text data into refined knowledge to facilitate research productivity. However, manual annotation and feature generation by medical experts are inefficient because they involve a complex process and require expensive and time-consuming labor.
* Our project is about Feature extraction from medical journals. It involves extracting meaningful pieces of knowledge related to clinical subjects like drug names, drug compositions, diseases, e.t.c from Medical Journals as unstructured data and unstructured data is computationally opaque they are converted to structured data and stored in csv file.
* We are building this custom tool which helps to minimize time commitments from domain experts and the manual efforts on researching content.



# PROJECT SCOPE

## Goals and Objectives

|  |  |
| --- | --- |
| **Goals** | **Objectives** |
| * Our major goal is recognizing Diseases, biomedical or Chemical composition names of drugs by named entity recognition (NER) which is an important task used to extract information and to recognize the Medical entities from Medical Journals. | * Maximize the number of named entities to be extract from the data. * Maximize the accuracy of the extracted entities. |

## Project Deliverables

|  |  |
| --- | --- |
| **Milestone** | **Deliverable** |
| * Understanding business objective and constraints | * Deliverable 1.1—Identifying Constraints and design the project architecture. * Deliverable 1.2— Data Extraction * Deliverable 1.3— Data Preparation |
| * EDA /Text preprocessing, NLP models | * Deliverable 2.1— EDA and Text Parsing * Deliverable 2.2— Model Building |
| * Model Building | * Deliverable 3.1— NLP model building using med7 library |
| * Preparing for Final Presentation | * Deliverable4.1 – show case and review * Deliverable4.2 – Final Presentation and documentation * Deliverable4.3 – Handover |

## Project Duration (start date: 01/11/2021 End date: 1/12/2021)

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Milestone** | **Date Estimate** | **Deliverable(s) Included** | **Confidence Level** |
| Understanding Business Objective and Constraints. | [01/11/2021]  -  [08/11/2021] | * Deliverable 1.1—Identifying Constraints and design the project architecture. * Deliverable 1.2—Explore PDFs to get insights on relevant data. | [High] |
| * Understanding data * Study on NLP models | [08/11/2021]  -  [15/11/2021] | * Deliverable 2.1— Explore various public forums to collect relevant data. * Deliverable 2.2— Data Preparation | [High] |
| * Model Building | [16/11/2021]  - | * Deliverable 3.1— NLP model building using med7 library | [High] |
| * PPT for Final Presentation and documentation | [23/11/2021]  -  [1/12/2021] | * Deliverable4.1 – show case and review * Deliverable4.2 – Final Presentation and documentation * Deliverable4.3 – Handover | [Medium] |



# PROJECT CONDITIONS

## Project Assumptions

* Work on data which is extracted from pdf format (medical journals)
* Create a .csv file from the extracted data.

## Project Issues

**Priority Criteria**

1. Error in selecting the named entity and time required.
2. Recognizing the medical terms accurately is a difficult task.
3. Issue in selecting library to extract medical entities.
4. Closed issue.

## Project Constraints

* Maximize the quantity of features extracted from the medical journals
* Minimize incorrect feature detection
* Minimize the run time of the model training

# Project Structure Approach

* Text Extraction in most efficient way
* Data Preparation and EDA
* Model development
* Model evaluation and fine tuning
* Documentation



# PROJECT REFERENCES

|  |  |
| --- | --- |
| **Milestone** | **Deliverable** |
| [ |  |
|  |  |
|  |  |

# APPROVALS

**Prepared by** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Manager

**Approved by** Sharat Chandra M\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Sponsor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Executive Sponsor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Client Sponsor



# APPENDICES

## Document Guidelines

## Project Charter Document Sections Omitted

