Project Initialization and Planning Phase

Date	Nov 2024
Team ID	Team-739662
Project Title	Chatbot based on Data Science Enquiry using NLP
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

The proposed solution involves developing a conversational AI chatbot that leverages natural language processing (NLP) and machine learning algorithms to provide data-driven insights and answers to users' queries. The chatbot will be integrated with a knowledge graph that stores data science concepts, relationships, and answers, allowing it to provide accurate and relevant responses.

Project Overview	
Objective	The primary objective of the chatbot is to provide accurate and relevant answers to users' data science-related queries, while also improving their understanding of complex data science concepts through intuitive explanations and visualizations.
Scope	The scope of the chatbot based on data science enquiry project using NLP encompasses designing and developing a conversational AI system that provides accurate and informative responses to users' data science-related queries. The chatbot will utilize NLP techniques to understand user intent and extract relevant information, and integrate with a knowledge graph to retrieve data science concepts, relationships, and answers.
Problem Statement	
Description	Design a conversational AI chatbot that can effectively understand and respond to data science-related queries, providing accurate and informative answers to bridge the knowledge gap and enhance the learning experience for individuals interested in data science.

Impact	The implementation of a chatbot based on data science enquiry project using NLP is expected to have a significant impact on the way individuals learn and interact with data science concepts, resulting in improved understanding, increased accessibility, and enhanced decision-making capabilities, ultimately leading to a more data-driven and informed community.
Proposed Solution	
Approach	The approach for the data science enquiry project using NLP involves a multi-step process. First, relevant data science-related texts, articles, and research papers are collected to build a comprehensive knowledge base. Next, the collected data is processed by tokenizing, stemming, and lemmatizing the text to prepare it for NLP analysis. An NLP model is then developed using techniques such as named entity recognition, part-of-speech tagging, and dependency parsing to extract relevant information from theprocessed data.
Key Features	**1. Conversational Interface:** A user-friendly conversational interface that allows users to ask data science-related questions in natural language. **2. Knowledge Graph Integration:** Integration with a knowledge graph that stores data science concepts, relationships, and answers to provide accurate and informative responses. **3. NLP-powered Query Understanding:** Utilization of NLP techniques to understand user queries, including entity recognition, intent identification, and sentiment analysis. **4. Data Visualization:** Capability to provide data visualizations, such as charts, graphs, and plots, to help users understand complex data science concepts. **5. Code Snippets and Examples:** Provision of code snippets and examples in popular programming languages, such as Python, R, and SQL, to illustrate data science concepts.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 8 GB

Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		
Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	e.g., tensorflow
Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
Data		
Data	Source, size, format	e.g., Kaggle dataset, 10,000 images