

Project Initialization and Planning Phase

Date	15 March 2024
Team ID	739675
Project Title	Drug classification using machine learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address the problem of accurately classifying drugs based on their properties using machine learning techniques. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements including hardware, software, and personnel.

Project Overview	
Objective	The primary objective of this project is to develop a machine learning model capable of accurately classifying drugs based on their chemical properties and biological effects.
Scope	The project will focus on collecting a comprehensive dataset of drug properties, preprocessing the data, developing and training a machine learning model, and evaluating its performance. The project will also explore the use of various algorithms and techniques to optimize the model's accuracy.
Problem Statement	
Description	The pharmaceutical industry faces significant challenges in classifying and predicting the effects of new drugs. Traditional methods are time-consuming and often inaccurate. An effective machine learning model can streamline the classification process and improve accuracy, thereby accelerating drug development and approval processes.
Impact	Solving this problem will lead to faster and more reliable drug classification, reducing the time and cost associated with drug development. This can ultimately lead to more effective treatments reaching patients sooner.
Proposed Solution	

Approach	The project will utilize a combination of supervised learning algorithms to classify drugs based on a dataset of known drug properties. The approach will involve data collection, preprocessing, feature extraction, model training, and evaluation. Techniques such as cross-validation, hyperparameter tuning, and ensemble methods will be employed to enhance model performance.
Key Features	<ol style="list-style-type: none"> 1.Utilization of various machine learning algorithms (e.g., Random Forest, Support Vector Machine, Neural Networks) 2. Comprehensive preprocessing pipeline to handle missing data, normalization, and feature engineering 3. Performance metrics to evaluate model accuracy, precision, recall, and F1 score 4. Exploration of advanced techniques such as deep learning and transfer learning for improved accuracy

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	2 x NVIDIA V100 GPUs
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, seaborn, matplotlib
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset