**ML Project**

* **Topic: Resume Screening for job profile**

Automating the resume screening process for various job profiles. Leveraging machine learning techniques, specifically K-Nearest Neighbours (KNN) classification and TF-IDF Vectorization, to efficiently categorize resumes into specific job roles. This project aims to enhance the efficiency and accuracy of the initial screening phase in recruitment, contributing to more effective and streamlined candidate selection processes.

* **Library Used:**
  + NumPy: for numerical operations.
  + panda: for data manipulation.
  + matplotlib: for 2D plotting of graph.
  + seaborn: for Statical data visualization based on matplotlib.
  + Regular Expression (re): for working with regular expression.
  + pickle: for serialization and deserialization python objects.
  + LabelEncoder from Scikit-Learn: A utility to convert categorical labels into numerical format.
  + TF-IDF Vectorizer from Scikit-Learn: A feature extraction method for text data.
  + Train-Test Split from Scikit-Learn: A function to split the dataset into training and testing sets.
  + KNeighborsClassifier from Scikit-Learn: A k-nearest neighbours classifier for classification tasks.
  + OneVsRestClassifier from Scikit-Learn: A strategy to extend binary classifiers to multi-class problems.
  + Accuracy Score from Scikit-Learn: A metric to evaluate the accuracy of the model.
* **Overview:**
  + Data Cleaning
  + Data Preprocessing
  + Feature Extraction
  + Train Test Split
  + Model Training
  + Model Evaluation
  + Model Serialization
* **Tools Used:**
  + **KNN Classifier:**

In our project, the KNN classifier is employed for multi-class classification during the resume screening process. This algorithm categorizes resumes based on their proximity to other resumes in the feature space, making it particularly effective for our diverse job profile classification.

* + **TF-IDF Vectorization:**

We utilize TF-IDF (Term Frequency-Inverse Document Frequency) Vectorization to convert textual data into numerical features. This technique reflects the importance of words in resumes by considering their frequency in individual documents against their occurrence in the entire dataset. TF-IDF Vectorization ensures our machine learning model captures meaningful patterns and relationships within the resume content, contributing to accurate job profile categorization.

* **Overview:**
  + Data Cleaning
  + Data Preprocessing
  + Feature Extraction
  + Train Test Split
  + Model Training
  + Model Evaluation
  + Model Serialization