# Metadata for Career Path Predictor

## Overview

This application predicts a user's ideal tech career path based on various personality traits and skills using machine learning. It processes user data, applies feature engineering, trains models, and provides career recommendations based on predictions.

## Technologies Used

* **Python Libraries**:
  + Pandas: Handles data manipulation and preprocessing.
  + NumPy: Supports numerical operations and matrix computations.
  + Seaborn & Matplotlib: Used for data visualization.
* **Machine Learning**:
  + Scikit-learn:
    - StandardScaler: Scales features for consistent model performance.
    - LabelEncoder: Converts categorical data into numerical form.
    - train\_test\_split: Splits data into training and testing sets.
    - PCA (Principal Component Analysis): Reduces dimensionality and improves efficiency.
    - RandomForestClassifier: An ensemble learning model using decision trees.
    - SVC (Support Vector Classifier): A classification model with kernel-based learning.
    - DecisionTreeClassifier: A simple tree-based classification model.
    - GaussianNB: A probabilistic classifier based on Bayes' theorem.
    - KNeighborsClassifier: A model based on nearest neighbors for classification.
    - StackingClassifier: Combines multiple models for better prediction accuracy.
* **XGBoost**: A gradient boosting library optimized for performance.
* **Joblib & Pickle**: Used for model serialization and storage.

## Key Features

* **Data Handling**:
  + Reads and pre processes career-related data.
  + Encodes categorical variables into numerical form.
  + Drops irrelevant features to improve model efficiency.
* **Feature Engineering**:
  + Normalizes data for optimal model performance.
  + Applies PCA to determine feature variance contribution.
* **Model Training & Evaluation**:
  + Implements multiple machine learning models.
  + Uses a stacking classifier to improve predictions.
  + Evaluates models using accuracy, precision, recall, and F1 score.
  + Visualizes performance using bar plots and heatmaps.
* **Result Interpretation**:
  + Predicts career paths based on user responses.
  + Provides insights through data visualizations.
* **Model Deployment**:
  + Saves trained models for future use.
  + Ensures efficient predictions for new users.

## Future Enhancements

* Improve model performance with hyperparameter tuning.
* Expand dataset to cover a wider range of career paths.
* Integrate real-world job market data for dynamic recommendations.
* Develop an interactive web-based interface for user accessibility.

This project leverages machine learning techniques to guide users toward suitable career paths based on their skills, personality traits, and interests.