**Dermatology Disease Diagnosis using Machine Learning**

**Overview**

This project aims to develop a machine learning model to assist dermatologists in diagnosing skin diseases based on patient symptoms. The model is trained on a dataset containing various features related to dermatological symptoms and corresponding disease classes.

**Contents**

**Data Preprocessing:**

* Basic Checks:
* Data Structure Investigation: Examined the structure of the dataset using data.info() to understand the types of features and presence of missing values.
* Handling Missing Values: Identified missing values, particularly in the 'Age' column, and replaced them with appropriate values.
* Data Type Conversion: Converted the 'Age' column from string to integer type for further analysis.

**Exploratory Data Analysis (EDA):**

* Visualizing Data Distributions: Utilized histograms and density plots to visualize the distributions of each feature in the dataset.
* Key Insights: Identified key insights such as the predominance of certain symptoms and the age range of the majority of patients.
* Boxplots and Countplots: Used boxplots to visualize the distribution of age and countplots to explore the distribution of age among different disease classes.
* FastEDA and Dabl: Leveraged libraries like FastEDA and Dabl for automated exploratory data analysis, providing quick insights into the dataset.

**Feature Selection**

* Identifying Relevant Features: Determined relevant features for modeling by selecting columns that were most likely to contribute to the prediction of disease classes.
* Excluded Target Variable: Removed the target variable ('class') from the feature set to ensure that the model is not trained on the variable it is predicting.

**Model Selection:**

Logistic Regression Model:

* Baseline Classification: Implemented logistic regression as a baseline classification model.
* Training and Evaluation: Trained the logistic regression model on the training data and evaluated its performance using accuracy, precision, recall, and F1-score.

Support Vector Machine (SVM) Model:

* Handling Class Imbalance: Addressed class imbalance using SMOTE (Synthetic Minority Over-sampling Technique) to oversample the minority classes.
* Training and Evaluation: Trained the SVM model on the resampled data and evaluated its performance, considering metrics such as accuracy, precision, recall, and F1-score.

Decision Tree Model:

* Tree-Based Classification: Utilized a decision tree classifier for non-linear decision boundaries and interpretability.
* Hyperparameter Tuning: Optimized hyperparameters of the decision tree model using techniques like grid search or randomized search.
* Training and Evaluation: Trained the decision tree model with optimized parameters and evaluated its performance on the test set.

**Evaluation:**

* Comparison of Models: Compared the performance of different models based on various evaluation metrics such as accuracy, precision, recall, and F1-score.
* Confusion Matrices: Analyzed confusion matrices to gain insights into the true positive, true negative, false positive, and false negative predictions of each model.
* Visualization of Results: Visualized evaluation results using seaborn's heatmap to enhance interpretability.

**Conclusion:**

Model Selection: Decision Tree Classifier emerged as the best-performing model with an accuracy of 97%, surpassing logistic regression and SVM.

Challenges Faced: Highlighted challenges encountered during the project, including data cleansing, understanding complex data relationships, and tuning hyperparameters.

Project Overview: Summarized the project's objectives, methodologies, and outcomes, emphasizing its potential impact on dermatological diagnosis and healthcare delivery.Feel free to incorporate these elaborations into your README file to provide a comprehensive overview of your project's data preprocessing, model selection, evaluation, and conclusion sections. Adjust the details as per the specifics of your analysis and findings.