**Documentation for Iris Flower Classification Analysis**

***Introduction***

The Iris dataset is a classic dataset in machine learning, containing measurements of sepal length, sepal width, petal length, and petal width for 150 iris flowers of three species: Iris setosa, Iris virginica, and Iris versicolor. The goal of this analysis was to explore patterns within the dataset and identify features that could be used to accurately classify the flower species.

***Exploratory Data Analysis (EDA)***

**Clustered Column Chart:**

A graph of different colored bars

Description automatically generatedA clustered column chart was created to visualize the median values of each feature for each species.

Key insights:

Petal length and petal width showed the most distinct differences between species, suggesting their potential as key features for classification.

Iris setosa had significantly smaller petal lengths and widths compared to the other two species.

Iris virginica generally had the largest petal lengths and widths.

**Stacked Area Chart:**

A screenshot of a computer

Description automatically generatedA stacked area chart was created to visualize the standard deviation of each feature within each species.

Key insights:

A large standard deviation in the petal\_length and sepal\_length for iris-virginica indicates that there is a lot of variances in the observed data around the mean. This indicates that the data observed is quite spread out. A small or low standard deviation is observed for iris-setosa which indicate that much of the data observed is clustered tightly around the mean.

**Key Insights and Patterns**

Petal length and petal width are the most promising features for distinguishing between iris species as can be seen by the following figure.

A screenshot of a computer screen

Description automatically generated

The petal\_length and petal\_width feature’s heatmap shows that the correlation between them is near 1 from which we can take them as a distinguishing feature.

A graph with different colored dots

Description automatically generated

Iris setosa is easily distinguishable from the other two species based on its smaller petal dimensions.

Iris virginica and Iris versicolor have more overlap in their features, suggesting potential challenges in their classification.

**Conclusion and Limitations**

EDA revealed patterns within the Iris dataset that can be used for flower classification.

Petal length and petal width are likely to be the most effective features for classification.

Further analysis, such as machine learning modeling, is required to evaluate the accuracy of classification based on these features.

**Limitations of the analysis include:**

The relatively small size of the dataset (150 samples).

The potential for biases in the collected data.

**Future Directions**

Build machine learning models to predict flower species based on the identified features.

Explore additional features or combinations of features that could improve classification accuracy.

Investigate potential biases in the dataset and their impact on classification.

Gather more data to strengthen the analysis and model performance.