

FINAL REPORT

Team: **Talismanic Vaders**

The Objective of this project was to predict the length of a product based on its product type using machine learning. The dataset provided contained information on various product types and their lengths, along with features like title, bullet points, and description. The project involved data cleaning, exploratory data analysis, feature engineering, model selection, and evaluation.

- The first step was to perform data cleaning, which involved checking for missing values and outliers. After checking for missing values, it was found that there were none. Outliers were identified using Z-score normalization, and data points with a Z-score of more than 3 were removed. This step was crucial to ensure that the model's predictions were not skewed by any outliers in the dataset.
- The next step was exploratory data analysis to gain insights into the data's distribution and identify any patterns. Box plots and violin plots were used to visualize the distribution of the data. A scatter plot was created to investigate the relationship between product type and length. From the exploratory data analysis, it was observed that the length of products varied significantly for different product types.
- Feature engineering was performed to prepare the data for model training. Columns such as title, bullet points, and description were removed from the dataset as they were not relevant to the prediction task. The final dataset contained only the product type and product length features.
- The deep neural network (DNN) model was built using TensorFlow and Keras libraries. The DNN consisted of 2 dense layers with 64 neurons each, followed by an output layer with one neuron to predict the product length. Mean absolute error loss function was used to train the model, and Adam optimizer was used to minimize the loss. The model was trained for 10 epochs, and its performance was evaluated on a validation set.
- The model was then used to make predictions on the test dataset. The predictions were saved in a CSV file and submitted. The final predictions had a mean absolute error of 0.57, indicating that the model was able to predict the product length accurately for the given test dataset.

In Conclusion, this project demonstrated the effectiveness of a deep neural network for predicting the length of a product based on its product type. The model's performance was evaluated using mean absolute error, which indicated that the model was able to make accurate prediction.