Abstract of Solution

**Introduction:**

As technology advances consumers expectations do as well. To meet those expectations, there is a need to implement new technologies that would elevate the value of a services. In turn this would ensure that consumers expectations are met. Guaranteeing that consumers are satisfied brings value as they are more likely to continue consume a product and would recommend that product to other individuals. This is a challenge that many companies have; this is why the investment in new technologies is so important.

**Problem:**

The problem that the Wells Fargo Campus Analytics Challenge 2022 presents is that of been able to offer costumers a feature that would increase their value in using Wells Fargo services. The way this would be done is by creating a model that would process a costumer’s transaction to categorize it into distinct categories. Making it easier for Wells Fargo customers to know where they are spending their money.

**Proposed Design:**

To solve this problem, I would like to use a machine learning model that allow the use of a training dataset with labels. The size of the dataset is also to be consider as it can have an impact on what model can be use. Familiarity to tools was also a factor when choosing a model. With all this in mind I selected the Sklearn LinearSVC model. This model can handle the amount of data that I have available and label data can be use. I have work previously with Sklearn making it the best option available.

The data that was give had a lot of useful variables that could be use on the model, but I only select a couple of them to ensure the best results. The once I selected where merchant\_cat\_code and coalesced\_brand. This where the ones I selected as they are the ones that provided the most description to what the transaction is about. coalesced\_brand give us an insight into the store, services, or brand that was purchased. merchant\_cat\_code gives a code that can be used to refences a dataset that enrich the description of the store, services, or brand that was purchased.

The way this problem is going to be solve is by importing the data that contains the transactions into our programming environment and manipulating said data to be use by the model that was selected. To complete this we need to look at the data and ensure that is process correctly. Once the data has been prep, we can move it to take features out of the transactions descriptions. This would allow me to pass it to the model so it can be used to be fitted. Once the model had been fitted a test dataset with transactions can be used to predict which category the transactions belong to. With all of that done we can export the final data set that contains the transactions and prediction into a csv file.

**Approach:**

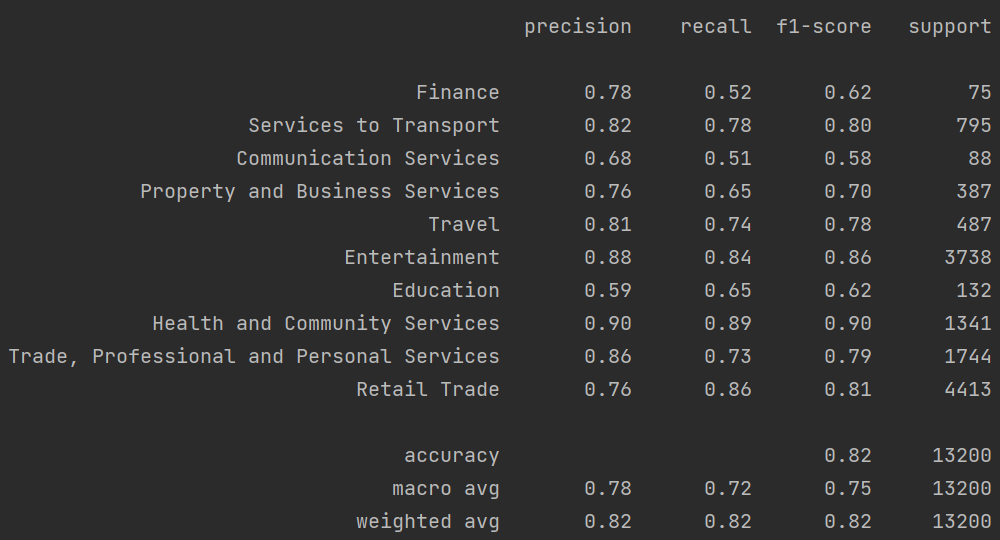
Once the data and model where selected we can start to approach the problem.

* Import Data
  + Use the pandas library to import data. The data to be imported is the training and test dataset provided on the challenges.
* Prep Data
  + Once the data is on our programming environment, we can prepare it. To prepare the data we combine all the text that would be use into a single column.
* Extract Features from Text
  + Use TfidfVectorizer form Sklearn to extract features from the text that is going to be used to fit the model.
* Model
  + Once the data is ready, we use it to fit the model with the given data.
  + Now that data has been fitted, we can use it to predict our test data.
* Export data
  + With the use of pandas we export the predicted data into a csv file for easy sharing.
* Extract Visuals
  + Use the training data to display visual that help us understand how our model works.

**Results:**

The picture below gives us a visual representation of how well our model can predict the categories of a transaction. As we can see the model is not 100% precent perfect as their outliers that are not label properly but they are minimal. ****

The following table allows us to see the numerical precision that the model has when predicting what type of category is going to be assign to a transaction.

****

The final data set is available in the following file final\_data\_CAC\_2022\_solution.csv

**Conclusion:**

The problem of been able to give customers satisfying and useful services as technology advances is here to stay and must be tackle head on. The Wells Fargo Campus Analytics Challenge 2022 gives us a challenge the correspondence with said problem. This is just a small sample of a real-world problem many engineers face. Yet it should be solved with the same tools and mindset as if it was an existent problem that a company is facing. That is why the use of machine learning was used to solve the challenge with the combination of industry standards.

**Visual Process Flow:**