How does the process actually work?

1. **Take our dataset then transform it to make it readable by our ML model**
   1. This is done by dropping columns that aren't readable or unnecessary in determining a grade for the venture.
   2. Also, by using standard scaler which allows us to take columns and change it into readable data by the ML model that were not readable before.
2. **Create a testing data set and a training data set**
   1. **Training data set**: contains 80% of the original data set and is fed to our ML model in order to allow it to see what KPIs correlate to what grades so that when it gets a new venture it can determine the grades without ever even seeing the venture before.
   2. **Testing data set**: contains 20% of the original data set however does not include any of the grades. This is so that the ML model can use what they gained from the training set analysis to now determine on its own the grade of each venture
3. **Next we send all this to our ML model for it to digest and figure out the testing set**
   1. SVM Classification Model: A classifier model that is able to determine complex relationships in datasets better than linear regression. It allows us use of multiple categories which gives us a letter grade scale to grade each venture's KPIs (A, B, C, D, F).
   2. Accuracy: This section in our program just tells us how many predictions the model was able to guess correctly but expressed as a percentage.
4. **Finally, we can go to the prediction stage**
   1. In the prediction stage all the user needs to do is enter the KPIs, Venture name, and date of analysis in order to determine by the model the probability of the given venture succeeding.
   2. Then the output will be your grade and that would wrap up the model.

**Example Use Cases of the Venture Selection Model in Predicting Future Outcomes:**

1. Citi Group: Was pretty good in the beginning (I'd give it about a high C or low B grade at the time the KPIs were taken) but is now struggling to maintain the metrics it had ten years ago. So given the metrics from approximately 10 years ago the model should predict that it should not be invested in.

<https://www.macrotrends.net/stocks/charts/C/citigroup/profit-margins>

Investment Name: Citi Group, Date: 9/30/2011, ROI: 1.36%, Debt To Equity Ratio: 2.35, Profit Margin: 3.03 , Social Media Engagement: 6 , PE Ratio: 5.6

A screenshot of a computer program

Description automatically generated

2.

3.

Explain the KPIs in the app and where all the KPI information came from:

**Venture Information on:** Investment Name, Free Cash Flow, Return on Investment, Debt-to-Equity ratio, and P/E ratio all were collected from macrotrends.net

**Venture Information on Date column:** Comes from the date in the charts where I recorded the KPIs for each venture. Intentionally I tried to get the earliest data I could on each venture in order to feed more relevant data when comparing big ventures to startups. The earliest the data goes back for most ventures was 12/31/2009 so if that was the earliest option that is what I chose for most of the ventures.

**Venture Information on Successful Column:** This column is primarily subjective because this is how we train the ML model to understand what we deem are good and bad businesses.

**Reasoning:** There's no universal rating based on how good or bad a business is so that's what we have to figure out ourselves. Thus, we create initial ratings to showcase what we believe the venture should be at, then the ML model determines based on scores we gave for each row of kpis (venture) what are the typical KPIs that align with each score. It should be noted that for determining successful scores that they weren't random numbers but educated decisions based off of metrics and articles found from credible sources to make our ML model predictions as accurate as possible.