Early Stage Project Success Measurement

By: Ahmed Hazmah Alfier Alshareef

Note: for more details check my well documented google colab notebook

1. Design

- **Why?**
  1. Rapid changes in technology
  2. Entrepreneurship is the mainstream
- **Goal:**

  Determine if an idea/project is worth pursuing or not (success or fail).

- **Who?**

  Two main category of beneficiaries would use this project:

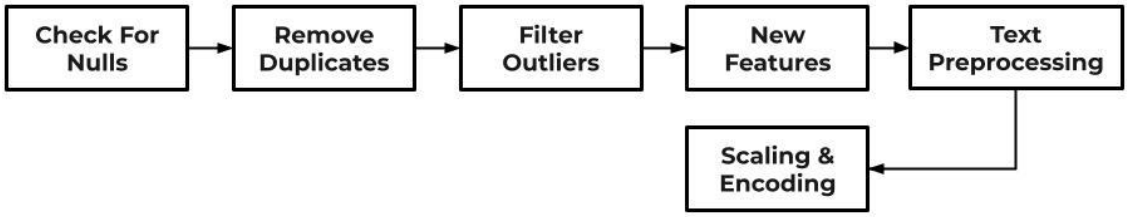
  1. **Entrepreneurs:** To assess the quality of their idea
  2. **Investors**: To determine what startups to invest in
- **How?**

  Measure how likely users are the users to pay for it. Success if the project achieve at least the financial goal or failure otherwise

2. Data & Algorithms

- **Data**
  1. Kickstarter dataset (Kaggel)
  2. Contains 13 columns
  3. 378,661 projects → After cleaning (124,235)

- **Preprocessing /Features Engineering:**
  - The Preprocessing pipeline



Metrics	Logistic Regression	SVM	Random Forest	Gradient Boosting	Bert + Gradient Boosting	Bi-LSTM	Bi-LSTM + NN	Bert
Accuracy	0.7445	0.7442	0.7447	0.7512	0.7505	0.7334	0.2666	0.7304
Precision	0.5664	0.5946	0.5539	0.5939	0.5905	0	0.2666	0.4838
Recall	0.1928	0.1375	0.2348	0.2210	0.2199	0	1	0.1647
F1	0.2877	0.2234	0.3298	0.3221	0.3204	-	-	-
AUC	0.7185	0.7184	0.7190	0.7384	0.7381	0.5	0.5	0.6218

3. Tools

- **Data Processing:**

  Pandas, and Numpy

- **Modelling:**

  SciKit-Learn, PyTorch, TensorFlow/Keras, and Pre-trained models (Bert & Glov)

- **Visualization:**

  Matplotlib, Seaborn, and Google Colab

4. Insights and Conclusion

- **Insights:**

1. **Model Range of Prediction:**

   (5,000 ≤ Goal ≤ 2,000,000)

2. **Best Dates:**
   - (Launch day: Tuesday)
   - (Launch month: October)
   - (Deadline day: Thursday)
3. **Best Categories:**
   - Music
   - Theater
4. **Worst Categories:**
   - Technology
   - Food
   - Film & Video

- **Prospective:**
    1. **Data is not sufficient:**
        - Bias models → more complex which needs more features e.g. Project description/Images
        - Unifying the currency of goal
    2. **Web presence:**
        - Integrated API / Stand alone website
    3. **Utilizing more GPUs & RAMs:**
        - Investigate more transformers/Pre-trained models