|  |  |
| --- | --- |
| Capstone Project Proposal |  |

*Ishika Jain*

**Business Goals**

|  |  |
| --- | --- |
| **Project Overview and Goal**  What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you’re labeling images, how will this help the business? | Project Name: Fraud Monitoring in Banks  The industry problem I am trying to solve is related to finance. AI and ML are able to provide unprecedented levels of automation that is ,it has faster data processing and less manual work. ML allows for creating algorithms that process large datasets with many variables and helps in finding correlations between user behavior and likelihood of fraudulent actions.  Examples:  Face Recognition is used to prevent credit card fraud. Cameras with face recognition can determine whether a credit card is in the hands of the rightful owner when buying at a physical point of sale.  Tracking suspicious IP addresses from which a financial transaction occurs may help prevent fraud with discount coupons as well as identify fraudulent intentions.  For instance, MasterCard integrated machine learning and AI to track and process such variables as transaction size, location, time, device, and purchase data. The system assesses account behavior in each operation and provides real-time judgment on whether a transaction is fraudulent. |
| **Business Case**  Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success. | Every year Banking Department loses millions of dollars in term of revenue that could have been saved if proper monitoring is done. Being indulge in Fraud related activities not only result in financial loses but also drop level of trust among customers. So by building this product manual detection of fraudster would be done using the software that has high accuracy and is more efficient. Research shows alone loan fraud can save $4687 per instance. There are other kind of fraud such as fake claim , tax related , mortgage related which can save millions of dollars. This will eventually improves Client experience , user data is used to make banks more productive and efficient with this AI helps banks by connecting data sets to make data driven decision for their clients. |
| **Application of ML/AI**  What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve? | There are different type of fraud that take place in banking systems.   * In credit card fraud detection ,cameras with face recognition can determine whether a credit card is in the hands of the rightful owner when buying at a physical point of sale. * Machine Learning has many algorithms that work with images and can classify them as fraudulent or not by finding out specific features and correlations.   For example, if we need to spot a fake watermark on the document with an algorithm, we should first train a model on a specific amount of fake and genuine documents so that it will easily discover a counterfeit one. The same rule applies to blurry digits or uneven lines that might be the result of an image- altering program such as Photoshop.  The business outcome I want to achieve is minimizing the fraud transaction so as to keep the hard earned money of hard working citizens safe. |

**Success Metrics**

|  |  |
| --- | --- |
| **Success Metrics**  What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison. | **Success metrics** are quantifiable data points that a business tracks and analyzes to gauge the **success** of its **product**. In this project(model) I will use non – fraud transaction as a metric for success. If there are no fraud transaction taking place in system then it is a good sign for success. We can easily measure this if no fraud complaint is registered within a particular span of time. For providing the baseline value to provide a point of comparison we will use stats before the model is deployed and stats after model the model is deployed. |

**Data**

|  |  |
| --- | --- |
| **Data Acquisition**  Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed? | There are number of well-known academic datasets available for this task for free. So, I would use them initially in prototyping. Since using only academic data might not be good as it might not represent the real-world scenario’s, so I would do some data collection myself and annotate using platform like figureEight, data from bank customers, investors, partners, and contractors can be used for different purposes, depending on which parameters are used to analyze them. This data is very crucial so we need to be careful and make sure that we are obeying all laws here .  We can collect the data as on-going process and keep refreshing model based on how well it is performing.  Since this data collection and annotations is not a cheap task as it can take few thousand dollars to much higher based on the amount of data you want to collect. |
| **Data Source**  Consider the size and source of your data; what biases are built into the data and how might the data be improved? | Source is a mix of academic dataset and personal collection and annotation. Usually dataset consist of bias on basis of gender, race and ethnicity in their model. For improving - this training data should appropriately represent and should be large enough to counteract common bias. Data used to train should be free from any kind of prejudices or false assumptions. |
| **Choice of Data Labels**  What labels did you decide to add to your data? And why did you decide on these labels versus any other option? | The three labels that are used for labeling the data is   * Non Fraud * Fraud * Unknown   These labels are easy to understand . Non fraud are those who is classified as non-fraudster by model.  Fraud Label is used by model to classify Fraud customer. “Unknown” to account for uncertainty.  The other way to do labeling can be like using Binary such as “0” for Non Fraud , “1” for Fraud and “Unknown” for Uncertain situation. This labeling is also correct but above gives quite a clear picture. |

**Model**

|  |  |
| --- | --- |
| **Model Building**  How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why? | Firstly I will be using Automated Tool one such is Google AutoML. By using this if I get satisfactory result what i wished to achieve then I will go to next step. So now I will consider the values of metrics that include Confusion matrix, value of precision and recall. If the model is achieving good performance then its done. Otherwise I would look for loopholes and try to improve as much as possible. If then also the performance is unsatisfied then I will use in-house team since automated ML tools are not performing well in all cases. |
| **Evaluating Results**  Which model performance metrics are appropriate to measure the success of your model? What level of performance is required? | Very common metrics like precision, recall, F1-score can be used. Usually we would want to detect all the fraudulent transactions in all transaction. So, you want this to be as high as possible like above 98.0 or 99.0 %. |

**Minimum Viable Product (MVP)**

|  |  |
| --- | --- |
| **Design**  What does your minimum viable product look like? Include sketches of your product. | AI In Banking For Fraud Detection Postings Ppt Powerpoint Presentation  Infographic Template | Presentation Graphics | Presentation PowerPoint  Example | Slide Templates |
| **Use Cases**  What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product? | This product is designed for Banking institution.  Scenario where this system is of use:   * Credit Card Fraud , there are numerous ways to secretly get credit card information. Use of pattern recognition and Data mining so to detect suspicious behavior to detect fraud. * AI and machine-learning tools can considerably reduce this number by enabling insurers to spot unusual patterns that a human might miss. For example, by comparing new claims to the existing data, [AI in the insurance sector](https://suyati.com/blog/how-ai-changing-the-face-of-the-insurance-industry/) can help detect claims that are unusually high and probably fake. |
| **Roll-out**  How will this be adopted? What does the go-to-market plan look like? | * **Prelaunch** – Conduct market research, test your product enough, prepared to fulfill orders, generate awareness and hype. * **Postlaunch** – Monitor your product performance and keep continuously improving, talk to customers and get their requests and roll out new features, fix bugs if there are any. |

**Post-MVP-Deployment**

|  |  |
| --- | --- |
| **Designing for Longevity**  How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product? | In long term, data evolves there will be new customer and new method of fraud transaction might be induced.  So to take all the situation into consideration we need to keep collection of more data and keep evolving our dataset so that model is able to deal with new situation it encounter with passage of time. We can train our model with new dataset using A/B testing. Here we use A/B testing(80% of traffic to old model and 20% to new model) . A/B testing (Split testing) helps to determine whether a new design brings improvement. |
| **Monitor Bias**  How do you plan to monitor or mitigate unwanted bias in your model? | Might be the model is performing well in some situation and able to correctly classify customer as fraud or non-fraud in mostly situation. But it lack in some situation so we need to add more in areas where are our model is not performing well and retrain it. We need to continuously monitor the areas where our model is not able to provide correct result for any reason may be due to algorithmic bias or due to data set. |