Capstone Project Proposal

<B.Akintade>

Proposal Intro: Focus industry for this Capstone Project Proposal is Banking/Financial Services. Havilah is a financial institution that offers a variety of Business Banking Solutions (BBS) to Business Clients (BC). The main BBS that Havilah offers are Cash Management Services (CMS) which includes Online Banking, One-off Disbursements CAD or USD, Wire Transfers, Electronic Funds Transfer (EFT), and Bulk E-transfers. Havilah adheres to the Anti-Money Laundering and Anti-Terrorist Financing Act and Financial Transactions and Reports Analysis Centre of Canada (FINTRAC) requirements. Havilah is committed to exceeding its Business Clients' financial needs by taking the stress out of banking and making banking seamlessly easy.

Business Goal	
Project Overview and Goal	The problem this proposal is trying to solve in the financial services industry is to reduce processing time it takes to enroll business clients for CMS, and make the service accessible and available to clients faster. The goal is to ensure that CMS for business banking is processed and delivered more efficiently within 24hours of business account opening as long as business meets all FINTRAC requirements. ML/AI is being proposed to solve this task because it will make CMS enrollment process faster, ensuring CMS needs are met by one touchpoint i.e. reduce number human of processing/touchpoints; and provide customer obsessed experiences that deliver value and make banking seamlessly easy.
Business Case	This is an important problem to solve because currently, Team Members (TM) of Havilah manually enroll BC accounts for CMS and this is a time-consuming task that has up to four touch points from the time a Small Business Manager confirms BC compliance for FINTRAC and gets BC consent for enrollment, sends requests to an Account Officer who downloads and completes the form, and then sends the form to CMS team via email attachment. It takes up to 3-5 business days to have this service finally completed and ready for BCs to access from their Online Banking platform. Building a product to solve this problem will make CMS enrollment faster; reduces manual processing; BCs have access to use CMS more quickly and frees up immediate cash transactions; increases Havilah's brand reputation among competitors; and drives more business for Havilah. Creation of automated processes will not only help replace manual, time-consuming tasks but will also add value to business client services by delivering faster enrollment for cash management services that business clients can have immediate access to.
Application of ML/AI	ML/AI will be used to automate the CMS enrollment process by creating an online enrollment form that lists available CMS. BC's desired CMS will be checked off and submitted online.

For this project

- Traditional Machine Learning approach will be used (which works better with smaller datasets) and not Deep Learning approach (which outperforms with large datasets)
- Input data is scanned images

To train/create a model that can correctly classify all the given types of cash management requests, and which will be able to classify any new requests into one of these classes;

Starting point in training the model is that it will need annotated/labelled data in form of images. Images are downloaded copies of requests that have been manually completed and scanned by team members. There're five different categories of classified requests, and the scanned training images/data is very representative of the real-world scenario. In order to create an algorithm to detect request types, the model will train on examples of previously completed requests.

1. A data labelling platform like Figure Eight will be employed to create the training data, upload data, design the annotation job, create test questions, categorize requests, and monitor annotation results. A set of instructions will be given to annotators which will include an overview, the steps that contributors need to take to complete the job, rules, and tips.
2. Neural networks are some of the most common models used in ML and for the purpose of this project, the focus will be on this specific type of model.
3. From my understanding of neural networks in this program, a single neural network is made up of numerous computational layers in which;

- Each of these layers will pass a decision to the subsequent layer until reaching the output layer, which will produce a final decision for the input data
- The data that goes into the network is a set of values
- These values could be pixels contained in previously scanned images

4. Neural networks are made up of neurons (aka nodes). One neuron is responsible for processing some input data and producing an output. For a neural network that processes visual data, such as the set of images previously mentioned, the inputs will be either spatial information or color information. When these color components or shapes are combined, which happens inside a neuron in the form of an equation (ex. 0.5red and 0.5blue = 1*purple), it produces an output signal that can help classify the initial input.

Ultimately, the neural network:

1. Breaks up the set of training data into a smaller, simpler model that is made of features
2. Learns how to combine the features and create thresholds/boundaries that can separate and classify any kind of data
3. Layers separation on top of separation layer to create more complex boundaries and group all kinds of data accordingly.

A model is made of visual features found in training data and a neural network combines these features, using a series of thresholds to identify any new data. As the machine learns by looking at many examples of given

	object/images/data, it will eventually learn that new pattern and data can be associated with a specific classification of requests submitted to enroll clients desired services.

Success Metrics

Success Metrics	• Faster online service enrollment instead of downloading forms to manually complete - since manual enrollment process is what Havilah has been doing over the years, I will make sure to collect specific number of CMS forms downloaded per week/month/year and compare this with number of online automated service enrolments • Immediate service access for business clients within 24hours instead of 3-5 business days - when CMS enrollment process is automated, there will be faster online enrollment confirmed by email notifying creation of additional tab created in BC's online platform for them to move funds as needed • Increase in number of happy and satisfied customers based on feedback received - feedback surveys sent out to BCs that have accessed new CMS tab to move funds; survey can be completed via pop up notification immediately BCs complete fund transfers or can be sent to their email • Free up valuable time for Small Business Managers to develop/have more quality time helping business clients - another custom survey can be completed internally by Havilah team members to provide feedback on how much time they are able to allocate to BCs in providing quality one-on-one services, and to also do other valuable tasks • Eliminates the need for a middle man (Account Officer) to process documents, which increases net revenue for Havilah – the Account Officer role will be made redundant when they do not receive any more requests from managers to download and manually complete CMS enrolment forms

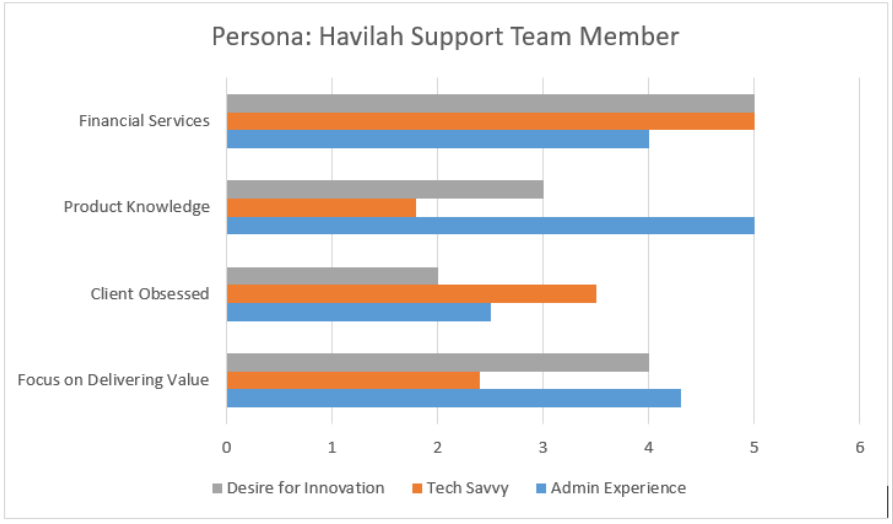
Data

Data Acquisition	• Data will be sourced from database of previously completed CMS requests submitted for manual processing in the last 12 months • Data acquisition is at no cost, though it is expected that five Havilah Team Members will be pulled from their current roles to dedicate time to pull records from data base of manually processed and completed CMS requests • Since sensitivity of data refers to the level of confidentiality the data is expected to have, sensitive PII including Business Client full name, driver's license, financial information, and other sensitive individual data is required. Protecting business client data and obtaining their consent will be a high priority in this process. • Data will become available on an ongoing basis since CMS enrolment requests are received weekly, so there will be no need to acquire a large batch of data

Data Source	Considering that hundreds of CMS requests are processed weekly/monthly by team members, biases built into the data are • Annotation bias due to human handling in the process of labelling and generating the training data • ML model bias due to bias generated by model outcome
	Traditional Machine Learning approach (which works better with smaller datasets) is what this model project is based on, and not a Deep Learning approach (which outperforms with large datasets). Some key points to consider when training a model, is to train it with all types of data that the model is likely to encounter in the future in order to build a robust model; and while there is no clear-cut rule as to how much data is required, it is generally advised to start with a few hundred examples of each target class and then scale up the amount of training data until a desired accuracy is reached'. On the other hand this project was looking at image processing through Deep Learning approach, the number of parameters in any CNN are very large and without sufficient amount of training data, the model will not learn anything; so the model may need up to 5,000 images to get a relatively solid confidence level and accuracy.
Choice of Data Labels	Data label examples are created to comply with CMS product names so that Havilah team members can easily classify and label CMS enrolment requests according to specific services requested by business clients. Data label names are; • *Havilah_Business_Solutions_Enrolment_CMS_Online Banking* • *Havilah_Business_Solutions_Enrolment_CMS_One-off_Disbursements* • *Havilah_Business_Solutions_Enrolment_CMS_Wire_Transfers* • *Havilah_Business_Solutions_Enrolment_CMS_Electronic_Funds_Transfer* • *Havilah_Business_Solutions_Enrolment_CMS_Bulk_E-transfers*
Proposed labeling scheme, and why the chosen labeling scheme was chosen. Strengths and weaknesses of such a labeling scheme; and potential weaknesses.	Some identified potential weaknesses of this scheme includes; > Human errors - since annotated data is needed to train the ML model, it will require steps including, uploading data, designing a job, creating test questions, launching and monitoring results. Labelling instructions will include an overview, steps that contributors need to take to complete the job, rules, and tips - during these processes, annotators can misunderstand instructions, miss a test question, and not be able to keep up with accuracy. What contributors miss the model may also miss. > Possibility of mislabeled, or dirty data which can significantly impact the model performance > Not having enough/robust data points and unevenly distributed data between different categories (of CMS requests) that the model is supposed to distinguish. This may result in model predictions that are biased towards classifying all data into one class, or predictions that have learned to find patterns that are irrelevant to the task at hand. > The model may not be able to generalize to best performance when faced with new user data if the training data used does not have many examples of previous different classes of data (in this case, the variety of cash management service requests)

Model	
Model Building	Havilah is an established financial institution with variety of IT transformation team, so model building will be resourced by leveraging on in-house experts - data, machine learning and software engineering team. Model training and/or hosting will be customized and built using an in-house team of experts this is because of the massive amount of business client data involved. Although expensive to get started, additional benefit with a customized modeling is that there will be unlimited use cases and Havilah will have total control over parameter tuning.
Evaluating Results	• Model accuracy - number of correct predictions the model got right • Execution time – amount of time it takes to complete the task/request • Recall - total relevant results correctly classified by the model • Precision - to measure exactly how much data is supposed to be categorized with a specific label, from all the test CMS examples assigned a label (EFT, Online Banking, One-off Disbursements etc). • F1 Score - this is used to find the balance between precision and recall, and used to measure the test's accuracy • Confusion matrix can also help to identify areas where model biases may occur and where the model is failing It is not expected that the model will be perfect and accurate to solve the business case problem at initial test stages; however it is expected to make enrollment process for cash management service faster and more effective. Also as a Product Manager, I will check with the science/engineering team to ensure that the new product works well with Havilah banking platform.

Minimum Viable Product (MVP)	
Design	MVP documents sketches of team member journey to enroll business client for required CMS service MVP sketch is attached as separate pdf document with Capstone project submission
Use Cases	Persona Example: Karen Knight • 32 years old • Recently moved in with boyfriend • No children • Diploma in Business Management, major in Financial Services • Earns ~$45,000 annually • Works out of a branch of Havilah but also has flexibility to work remotely from home • Obsessed with delivering value to clients Karen's Pain Points • Frustrated by manual processing of requests • Irritated by poor service delivery • Concerned that robots might take her job • Concerned about the economy and job options



Use Case:
To reduce manual CMS processing and ensure that CMS for business banking is processed and delivered faster and efficiently within 24hours; to make cash flow management seamlessly easy for business client banking needs

Users will access product with their Havilah work email and official employee number. Designated technical team
will be available to support team members via chat room or phone call if need be.

Roll-out Adoption and go-to Market Plan

Product will be adopted by having different teams try out versions of the new product/tool which will be made available for different service enrollments e.g Online Banking Enrollment, Electronic Funds Transfer (EFT) Services etc. Feedback will be encouraged from all team members trying out the product.

As a Product Manager, I will:
- Check with engineering team to ensure data, operational and security compliance with Havilah banking platform
- Work with product marketing team to ensure that announcements Team Leaders are aware of upcoming product roll-out so that they can in turn let their various team members know; I will also ensure that the marketing team post expected roll-out dates on internal online platform one month, two weeks and 48 hours ahead of first MVP roll out, so we can get the product into the hands of users as quickly and fast as possible
- Metrics collected from user tests will also help with product marketing
- End-user business clients will be notified to expect changes on their online platforms that will make cash management/fund transfers easier

Post-MVP-Deployment	
Designing for Longevity	Product will be improved in the long-term by: • Using product feedback/new information from users and applying the feedback to new product versions being released • Continuous repeat of prototyping and iterating process quickly and faster with new user information • Refreshing and updating the model with new data to prevent model staleness where predictive power of the ML model decreases over time as strengths or user state changes In addition to preparing for long time, when the model sees new requests submitted that it hasn't seen before, or it has low confidence in, the best practice that Havilah will put in place is to incorporate real humans (team members) into the training pipeline so that the natural language processing system/model can route unrecognized data/image to a live support system for resolution i.e. the model sends the non-confident data to a team member who can look at it, correct the classification and then feed that information back to the model. Typically this is where versioning comes in or continuous learning, and the model can be updated or improved based on new information, and the next time it sees a particular request e.g. wire transfer enrollment, it will have a higher confidence level than before, in which CMS request submitted for wire transfer enrollment will not be processed the same way as CMS request submitted for Electronic Funds Transfer. When the model learns from real data, it builds confidence and ensures consistent improvement over time. Real-world data is different from the training data in the sense that; > Training data is used when a model is learning i.e this is the data used by a model to find patterns that distinguish different classes of images, and the model ultimately learns from these images > Real-world data (RWD) is data/information obtained from various real-world sources that are associated with outcomes in order to have an accurate representation of what was done to guide clinical decisions, from diagnosis to treatment*. RWD are collected, de-identified and stored in a variety of sources to be later analyzed alongside similar data; and may provide new insights about a medication beyond what is available from a clinical trial alone*. Product can learn from new data by: > Setting up the model in a dynamic way that it is continuously trained from new input > Updating the data to include more relevant examples for model to gain expertise > Getting rid of incomplete and inaccurate data > Ensure that training and test data are balanced

A/B Testing	A/B testing helps to make more data-driven decisions, and it will be employed to improve product by: > Deploying two different models e.g. deployment can initially be for two CMS service enrollments - Wire Transfers and Electronic Funds Transfer (EFT) > Splitting acquired data into two models in order to track performance metrics associated to each model; after which a decision can be made on the winning model and/or if the lowest performing model can be replaced with the better performing winning model > Continuously run tests on statistically significant sample size until reliable results are obtained from the model > Consider the potential cost of implementing the new model/product as there may be high cost to launch it
Monitor Bias	Monitoring and mitigating unwanted bias in the model will be achieved by; > Defining the business goals and success metric from the very start of the project > Making sure that the specified success metrics are continuously monitored and corrected > Continuously tracking the model output for accuracy, execution time etc., in order to drive the business outcome

References:
https://www.pharmaceutical-technology.com/features/real-world-data-improving-outcomes/
https://www.mdedge.com/clinicalneurologynews/article/177311/business-medicine/real-world-data-machine-learning-and