**AmEx Ignite Proposal**

**Section A: Personal Details**

Please provide details of prior background in machine learning/artificial intelligence (including courses/projects/competitions undertaken).

**Experience**

**Data Science Associate*- [Publicis Sapient, Bengaluru]***

5th May – 5th July, 2019

• Offered with **a Pre Placement Offer** to join the company full time as **Senior Associate Data Science** after graduation.

• Designed a **Flask API** which analyses the threat percentage of YouTube channel based on **text, image and video features**

• Used **NLP** on textual data of YouTube, used **Hierarchical Attention Network** to predict the threat percentage of the videos

• Built a **scraper** to extract YouTube data. Used **Feature extractor** to extract frame level and video level data from videos.

• Used various **deep learning** models to predict which videos will be best to run **YouTube ads** for best returns.

**Data Analyst *- [Medbay, Noida]***

1st Dec - 31st Dec 2018

• Used **IIB data** to create models to **predict** the hospitalization rate, claims rate and claim **amount** of potential customers.

• Predicted the price of Health Loan Card for population categories and analyzed **the financial cost of risk and uncertainty**.

• Gathered, understood and documented detailed business data, **reports** and **visualization** for Investors and internal use.

**Cryptocurrency and ICO Investment Analyst *- [Tropyc, Bengaluru]***

15th May - 15th July 2018

• Improvement of metrics for investment **rating algorithm** for a diversified billion-dollar crypto currency and ICO market.

• **Extracted, filtered, cleaned, and visualized** ICO data and key performance indicators to interpret trends and patterns.

• Built the **analytics page** based on data-driven understanding of ICO, its valuation, market size, and other key indices.

• Created **1,000,000,000** decentralized ERC20 utility token "**Tropycoin**" on Solidity which is powered by **Ethereum**.

**PROJECTS**

**Cognitive Workload Detection using EEG data and Machine Learning *– [Prof. D Samanta , IIT Kharagpur]***

August 2018- April 2019

• Formulated **Brain-Computer-Interface** solution for measuring **Cognitive Workload** for classifying individual and task

• Designed experiment for the collection of **EEG data** in industrial scenarios depicting various levels of cognitive workload

• Performed Signal Processing & Feature Engineering on EEG data using channel, **Feature Extraction** & Optimization.

• Modeled for Cognitive workload measurement using K-NN, Random Forest, Decision Tree Classifier, SVM and MLP.

**Autonomous Vehicle Optimization – [Prof. Pallab Dasgupta, IIT Kharagpur]**

November – December, 2017

• Designed an end to end **Feed-Forward Neural Network** architecture for training the ML Algorithm for **autonomous driving**.

• Trained the architecture by converting colored image input to grayscale using **Image Processing** and Pattern Recognition.

• Boosted performance of the learning algorithm by improving the dataset with enhanced contour edges for less error.

• Built up the network to produce correct steering direction of the automobile using Back-Propagation Algorithm.

**COURSES**

Machine Learning | StanfordUniversity | Andrew NG

Introduction to Data Science in Python | University of Michigan | Kevyn Collins Thompson

Applied Machine Learning in Python | University of Michigan | Kevyn Collins Thompson

Soft Computinhg | IIT Kharagpur | Dilip Kumar Pratihar

Educational Data Analytics | IIT Kharagpur | Jiaul Hoque Paik

**Section B: Proposal**

* Describe the project you wish to work on. How did you come up with the idea?

I propose to design, develop and deploy a Proof of Concept of Federated Learning on local server, raspberry pi and android devices. It will be intended to achieve decentralized, distributed, on-device machine learning. This prototype could be upgraded to use the processing power of millions of devices. It will do so without sharing their private data of clients /devices to central server. This will not only allow for privacy-preserving analytics but also the harness the processing power of devices which would be enormous.

The prototype can be used for several business use case like : training of video/image data on individual cars for autonomous driving. This will avoid the need for accumulating a humungous dataset on a centralised cloud server for training and will even reduce the cost for enormous processing power, since the models could be trained in the car's chipset itsef.

-Using federated learning to understand the purchase / card-swiping behaviour of individual customers without breaching their privacy and giving them personalised offers based on their purchase behaviour. Even more - the data of other companies can be combined without leaking the private data and merged to get a better understanding of customers and hence target the customers for unified profitability of the data providing companies.

-Building of personalised recommendation systems for users without data privacy issues since the data will remain in the hand of respective users.

-Development of a secure federated learning system, which will enable collaborators to train a shared machine learning model for healthcare without exchanging confidential patient data thus helping the healthcare sector to remove the barrier of confidentiality issues.

Idea

Once I had to train a huge embedding corpus for a freelancing project. I needed a GPU for doing that, my laptop didnt support that computation. I started looking up for ways where i can train such a huge amount of data and in less time. I went through several areticles and blogs but didnt find much. Although at the end I used google colab and it worked fine.

But this left me thinking about the large data problem.

Since 90% of todays data has been created in last 2 years and in next 5 years the systems would be genearting so much of data that it would become almost impossible to get analytic solutions out of them and also the computing power and computing time will become costly to run on such extensive dataset. So I started looking for answers for this particular problem. Talked to some professors and did some digging.

I came across federated learning and i found that it not only solved the problem of extensive data training but also it took care of privacy issues. And since then I wanted to work in this domain and I have been researching more about it. This technology is relatively new and it has the potential to disrupt how we train models and how data privacy can be taken care of.

Facebook had to pay a fine of 5 Billion dollars due to privacy leak after the cambridge analytica scam. Google also had to pay 50 million pounds for GDPR violations. Federated learning can not only solve these issues but can give us access to enormous data which will be so much valuable for building extensive models and can solve so much of business problems at hand.

Hence I wanted to work on this and make a prototype on local servers/ raspberry pi or android and make use of the technology for creating a handy application based product which could be readily used to solve business use cases.

* Provide a detailed project plan with project milestones and estimated timelines.
* Are you starting this project from scratch or is it an on-going project? If it is a work in progress, provide more details.

I will be going to start the work from scratch. But only a few works and research have been done on this particular technology. And I would be referencing them for my project. The technology is still in very novice phsae and using Federated learning as an application for training on decentralised devices for various use case is something that has not seen much of usage.

My objective is to make the technolgy handy for direct application in solving real world business cases.

* Would you require financial funding for access to new data sources or computation? If yes, please elaborate on the requirement clearly stating the purpose.

For building the federated learning prototype on Raspberry pi, I would need some financial support to buy them. For the building the proof of concept, 2 raspberry pi would be enough for the deployment and testing. Hence a financial aid of 6000-7000 INR would be sufficient to carry out the whole project.

In addition to for training the dataset. I can use publically availiable data for the training, testing and deployment.

* Describe the techniques you intend to explore to accomplish your project. Also, specify details of how you plan to obtain data to train/test the algorithm.

Pysyft and tensorflow federated.

* What kind of help do you expect from the Mentor who will be guiding you? (The mentor will be an experienced data science expert, working at AmEx)

guide me through the overall process. With the expertise how can i make use of the technology and deploy it for even more concerned use cases

* What should be the success metrics on which your project should be evaluated?

if i am able to create a protoptye applying federated learning on raspberr