[redacted]

[redacted]

EDUCATION

The University of Texas at Austin, *Master of Science, Business Analytics (GPA: 3.9/4.0)* May 2023

- Coursework : Natural Language Processing, Deep Learning, Statistical Models, Feature Selection & Extraction, Cluster Analysis, Forecasting, Predictive Modeling, Optimization , Unsupervised Learning, Data Visualization, Marketing Analytics, Finance

University of Pune, *Bachelor of Engineering, Electronics and Telecommunication (GPA: 4.0/4.0)* June 2020

PROFESSIONAL EXPERIENCE

2Block Analytics **– *Machine Learning Engineer (NLP) Intern*** Jan 2023 – May 2023

- Built 91% accurate classification models using GPT-3 API to streamline business document processing, saving 10000 dollars
- . Collaborated with start-up founders to develop text analytics product and deploy it on AWS EC2 instances
- Finetuned and trained LLMs like BERT and GPT-3 for efficient filling of due diligence questionnaire (DDQ) form
- Upgraded MPNet and GUSE LLMs to get embeddings of similar text pairs, achieving average cosine similarity of 98%

[redacted]aya **– *Machine Learning Engineer*** July 2020 - June 2022

- Spearheaded the development of a conversational intelligence AI product based on 10 KPIs using LLMs and NLTK to calculate the Response Quality Score (RQS) of the conversation between a customer and AI chatbot
- Led the development and deployment of AI based Contact Center to improve customer interactions by optimally pairing customers with contact center agents, boosting average revenue per customer contact by 1.5 times
- Implemented BigQuery on Google Cloud Platform, causing a 50% reduction in data processing time
- Designed ETL pipelines in AWS Glue to migrate customer data from S3/Text files into Redshift
- Achieved 75% accuracy in predicting customer churn using stacked ML model of XBGBoost, LightGBM, and RandomForest
- Deployed data-driven ML model on AWS Sagemaker with data stored in AWS S3, achieving 20% faster training time
- Handled data imbalance using SMOTE, boosting model performance by 20% in F1-score & reducing false negatives by 15%
- Designed and executed A/B tests to determine product KPI performance across different variations of features
- Administrated end-to-end MLOps lifecycle to develop and deploy end-to-end ML solutions for 3 clients
- Developed a scalable and flexible framework in PySpark to segment the customers using K means clustering
- Wrote 100+ SQL custom functions to obtain insights from multiple tables of PostgreSQL storing 1B+ records
- Created Tableau dashboards to narrate customer acquisition recommendations, achieving 10% increase in retention rate

TECHNICAL SKILLS

- **Programming Languages**: Python, SQL, R, MATLAB, C, C++, HTML, CSS, BASH(Linux Shell Scripting)
- **Libraries**: TensorFlow, PyTorch ,Gurobi, Scikit-Learn, Pandas, NumPy, Matplotlib, Spacy, NLTK, Keras, Hugging Face, Open AI
- **Cloud Frameworks**: AWS (S3, EC2, Glue, SageMaker ,Athena, Lambda, Redshift), Azure(ML Studio), GCP (BigQuery)
- **Tools**: Power BI, Tableau, SCRUM, Agile, Git, Docker, JIRA, Alteryx, Google Analytics, SAS, Microsoft office suit
- **Databases and NoSQL**: Databricks, PySpark(MLLib), MySQL, MongoDB, Postgres, MS SQL
- **Concepts**: MLOps, Data Structures and Algorithms, Object Oriented Programming, A/B testing

PROJECTS

Optimization, Integer Programming *(Gurobi,Tensorflow,numpy)*

- Created a portfolio out of NASDAQ-100 stocks using 2019 and 2020 data to get maximal returns using integer programming
- Found the optimal number of stocks to include in the portfolio by accounting in diminishing returns

Optimization of CNN machine learning algorithm *(CNNs, Deep Learning, Optimization)*

- Designed an optimizer to reduce the time complexities of CNNs, enhancing accuracy of model from 89% to 94%
- The proposed optimizer outperformed SGD and Adagrad optimizers with 20% faster training time.

Information Management for Amazon Ecommerce *(SQL, AWS-S3, MongoDB, Data Lake)*

- Designed three distinct transaction models for Amazon Ecommerce and generated 40 GB of synthetic data
- Implemented data warehouse DDL and ETL processes to load transactional data into the data warehouse
- Assembled two data lakes using MongoDB and AWS S3 services, and developed Python and JSON files for data visualization

Food to Recipe/Nutritional Information Generator *(Transfer Learning, Deep Learning, Text analytics)*

- Web scraped, pre-processed, and resized 166,000 food images from food.com along with their recipes
- Trained VGG-16 and ResNet CNNs on scraped data to predict the recipe of an input image, achieving Jaccard distance of 0.78