



TEK07217 (Data Scientist - 9 yrs)

■ BSc

Highlights

- **Bachelor of Science in Mathematics** with **9+ years** of experience in Data Science building end to end data pipe-line, Machine Learning solutions, Artificial Intelligence, Natural Language Processing, Computer Vision, Data Analysis & Insight Automation, data processing & data mining algorithms to solve challenging business problems in multiple domains i.e. Banking, Finance, IT, Airlines Transportation & Telecom.
- Presently working as a Data Science Consultant (Individual Role-Freelance) with one of AWS partner and developing AI & Machine Learning Models, automated Data pipelines & Data Analytics solutions using latest techniques & tools – Python & its libraries, AI, Deep Learning, NLP, LLM, SQL & NoSQL databases, Data Studio, Rest APIs, AWS, Google Cloud Platform.
- Expert in implementing advance Machine Learning & Deep Learning techniques using Python & its libraries and cloud services to solve various Business problems i.e. Image Classification, Object Detection, Predictive Models, NLP chatbot, Large Language Models, automated KYC, Text Classification, Text Extraction, ML-Cross-sell Campaign, Credit Risk, Identity theft etc.
- Expert in building & optimising end to end automated large structured and unstructured data pipelines needed for Insight automation, ML & AI model building, auto training, auto tuning and auto deployment on Google AI Platform and AWS Sage Maker/ECR/Lamda.
- Always curious to explore and research innovative techniques to optimise ML, AI & Data analysis solutions for higher accuracy and specific business purpose/domain.
- Expert in implementing modern methodologies for data analysis, Image Processing, NLP, ML and AI solutions.
- Expert in analysing large structured & unstructured data to produce actionable insights, process Optimization and building machine learning, AI, NLP & Deep Learning models using latest tools & techniques which helps business to improve customer experience and revenue with mathematical quantification of positive & negative impact.
- Quick learner, willingness to work hard and have ability to take initiative with a creative blend of mind and deliver projects with/without any supervision in a very fast-paced environment .
- Effective communicator with excellent numerical abilities, ability to work under pressure. Possess a flexible & detail-oriented attitude.

Skills

Primary Skills

- Artificial Intelligence
- Machine Learning
- AWS SageMaker
- Natural Language Processing
- Keras

Secondary Skills

- Anaconda
- Python
- Jupyter
- Pytorch
- Numpy
- Scikit

- Spacy
- Tensorflow

Other Skills

- Expert in Data Mining, Data Management, Machine Learning, Predictive Modelling, Artificial Intelligence, Deep Learning, Computer Vision-Image Classification & object detection, Natural Language Processing, Transformers, Large Language Model, OCR, Data Processing & Cleansing, Data Analytics, ML Model Evaluation, Report Automation, Quantitative & Qualitative Analysis, Process Optimization, Business Intelligence, Data Visualization.
- Expert in developing ML & AI solutions, Data Analysis, Insight automation and Process optimization.
- Expert in Python Programming & ML libraries, SQL & NoSQL databases, Teradata, Google Data Studio, Tableau, Cloud Computing-GCP&AWS, Computer Vision.
- API Development to integrate Models with websites, Internal applications & Databases.
- Google Cloud Platform: AI Platform – Python Notebook instances, Cloud Storage, Big Query, SQL, firestore, Compute Engine, App Engine, IAM, Cloud Run, Cloud Function, Workflow, Docker Registry.
- AWS – S3, Redshift, Lambda Function, Step function, Amazon Gateway API, Textract and Sage Maker
- Visualisation: Tableau, Power BI & Google Data Studio

ML TECHNIQUES & Tools

- Regression, Classification, Image Classification, Object Detection, Text Mining, Text Classification, Text Summarization, Name Entity Recognition, Merchant Recommendation, Sentiment Analysis, Cross-validation, Pre-processing, GridSearchCV, Feature Extraction, Feature Engineering, Corpus, Bagging-Boosting, Collaborative Filtering, Market Basket Analysis, Segmentation, Text Extraction, LLM.
- Logistic Regression, Decision Trees, Random Forest, XGBoosting, K -Nearest Neighbors, SVM, Naïve Bayes, K-means, Ensemble Models, Apriori, Cosine Similarity, Time Series Forecasting, Deep Learning, Multilayer Neural Network, CNN, LSTM, RNN, Transformers, GANs.
- Numpy, Pandas, Seaborn, Matplotlib, Scikitplot, Scikit learn, Keras, Tensorflow, Pytorch, Word2vec, NLTK, Word cloud, Spacy, Confusion Metric, Sensitivity, Specificity, Precision Recall, F1 Score, Anaconda, IPython, Jupyter Notebook, Google AI Platform, AWS Sage Maker.

Projects

Project 1 - IT Industry (39 months)

- **Problem Statement:** Build AI Model which can answer customer questions on sports
- **ML-Models:** Pytorch, NLTK–tf-idf vectorizer, word2vec, LSTM, Transformer, stop word, sent_tokenize, word_tokenize
- **Solution:** Improved customer experience by handling complaints on priority within timeline and reduce to maintain human resources.

Responsibilities:

- Individually I am responsible for developing end to end Advanced analytics solutions which includes developing automated data pipeline, Predictive Model development for sports, Text Classification, Data mining, Data Management, communicating with management and internal web development team on the project progress.
- Develop automated solutions and streamline data collection, data cleaning and generate insights for business decisions.
- Design, develop & deploy a Language Model on AWS which can answer customer questions on sports.
- Develop automated ML pipeline for auto training, tuning and deployment.
- Leverage Chat-GPT for sports insight generation and prediction.
- Generate ML performance insights using automated solutions.

Tools Used:

- Python, Pandas, Numpy, Matplotlib, Scikit-learn, Pytorch, Keras, Tensorflow, NLTK, Spacy, SQL, Google Data Studio, Amazon Web Services, Docker, Lambda, ECR, Step function, Sage Maker, Chat-GPT.

Project 2 - IT Industry (12 months)

- **Problem Statement:** Recommend the most appropriate merchant based on customer behaviour & spending pattern.
- **ML-Model:** Text Mining, TF-IDF Vectorizer, N-grams, Cosine similarity, Item based Collaborative Filtering (Predicted using AWS Sage maker for better computational process)
- **Solution:** Using Item based collaborative Filtering Recommended the Merchants based on customer spending behaviour.

Responsibilities:

- Responsible for implementing modern Advanced analytics solutions which includes develop data pipeline, Artificial Intelligence & Predictive Models, Email & chat classification NLP solutions, Insight & Dashboard Automation, Data mining, Data Management, Quantitative & Qualitative analytics, communicating with management and business team along with internal analytics team on the project progress.
- Designing & automating data analysis reports & Dashboards for better performance tracking and consultation.
- Developed & optimised end to end large structured and unstructured data pipelines needed for Insights, model building, auto training, tuning and auto deployment on cloud.
- Design, develop & deploy ML models to run targeted campaigns for different sport products.
- Design, develop & deploy Natural Language Processing model for email classification & text summarization.
- Develop an Object detection model to process customer selfies and get insights.
- High level and deep data analytics interpretation for management business decisions.
- Campaign management, performance tracking and insight generation on campaign success.
- Develop churn models to enhance customer engagement.

Tools Used

- Python, Pandas, Numpy, Matplotlib, Scikit-learn, Pytorch, Keras, Tensorflow, NLTK, Spacy, SQL, GCP(AI Platform, Google AutoML, Docker Registry, Cloud Storage, Firestore, Big Query, Computer Engine, Cloud Function, Cloud Run, Workflow, Cloud Scheduler), Power BI, OLAP, Datorama.

Project 3 - IT Industry (48 months)

• Problem Statement:

1. Find high propensity customers who will buy Personal Loan in near feature for Marketing Campaigns.
2. Find high propensity to cross sell Credit Card to CASA Customers to increase customer base and customer experience.
3. Find high propensity customers who will become Revolver in near feature.
4. Developed NLP language model (Chat Bot) to answer customer questions related to Banking products and help training the bot systematically(Base Algorithm - IBM Watson).
5. Develop Image Classification and Object Detection model process customer selfie(with their house and office in background) and use the output to assess risk.

• ML-Tech:

1. Extracting Raw data from source (Data Base) Data Mining, Data Cleansing, Data Wrangling, Feature Selection, Feature Engineering, Statistical hypothesis testing, Sampling for Imbalanced Data, Multilayer Neural Network, Evaluating Bias Variance, Error Analysis, back testing, Score New Data.
2. Ensemble Models
3. Neural Network
4. Natural Language Processing, td-idf vectorizer, cosine similarity and Truncated SVD.
5. Fast R-CNN, Interpolation, Decimation, Min-max Normalization, and data augmentation techniques.

• **Solution:**

1. Neural Network model have been giving the 5% better lift than manual campaigns. Quantitative.
2. Neural Network model have been giving the 3.8% better lift than manual campaigns.
3. Neural Network model build to predict in advance who will become revolver so Bank can provide offers in advance.
4. Humans not needed to answer all customer calls/emails/chats and gives ability to customer to get answers to their queries in realtime.
5. Detected objects from customer selfie and tagged with other customer data to assess risk better.

Responsibilities:

- Individually I am responsible to deliver end to end Advanced analytics solutions which includes AI & Predictive Model development, Natural Language Processing, Computer Vision, Data mining, Data Management, Quantitative & Qualitative analytics, communicating with management and business team along with internal analytics team on the project progress & impact on business in respective of revenue and customer experience.
- Developed & optimised end to end large structured and unstructured data pipelines needed for Insights, model building, auto training, tuning and auto deployment on cloud.
- Design Metrics & statistical reports to measure digitization of all Banking products.
- Find business opportunities and lacking areas by analyzing large structured & unstructured datasets (Transactions, Demographics, Chats, Emails, Social Media, Website) and built predictive models/NLP/machine learning/Deep Learning solutions.
- Propose and work with business teams to implement solutions of Banking problems using state of the art Data Science techniques – Predictive Modeling, Machine Learning, Deep Learning, NLP.
- Responsible for End to End Predictive Modelling process and implement large scale solutions using Python Programing, Teradata, Base SAS & SAS E-miner, Cloud Computing.
- Work with internal team member to understand the current solutions and how to improve & automate them using latest data analytics, ML and Deep Learning techniques.
- Analyze large data sets from different sources, produce actionable insights and create automated Dashboards in Tableau & Data Studio.
- Provide data driven insights & recommendations to strategic business questions.
- Develop, document, implement & maintain BI reporting solutions for senior management & other internal teams on Predictive Models output.
- Co-ordinate with data engineers to fine tune ETL process to collect useful information in data bases from source systems which can enhance analysis and help to get more insights on customer behavior.
- Responsible for technical definition & consulting assistance to facilitate statistical data analysis, data integration & management and preparation of reports to support business specific deliverables.
- Built a predictive model to identify Customers without PL has high propensity to buy PL (CASA & CC Base).
- Built a predictive model to identify CC Customers who are going to default.
- Built a predictive model to identify CASA customers without Credit Card have high propensity to buy Credit Card.
- Built a Recommendation System to recommend merchants by understanding customers implicit & explicit interests by analyzing their historical transactional behavior.
- Sentiment Analysis-Build a NLP model to evaluate customer's feedback from multiple sources (Internet Banking, Social Media (Facebook, Twitter, Chat bot) & Call Centre.
- Model is back tested by real time manual campaigns and compares the results of both (Response Rate of Manual Leads vs. Model Leads).
- Experience in Personal Loan, Mortgage Loan, Credit Card, CASA etc.
- Hands on experience on Attributes Selection for ML & Evaluating the Models.
- Build email and text classification NLP solution for prioritised action.
- Develop NLP language model and train on all Banking products, create auto training, tuning, deployment pipeline on GCP-AI platform and integrate with Bank's website to answer customer questions.
- Closely working with GM/COO to provide day to day data analysis which helps management in strategic decisions.
- Streamline KPI reports for all the Banking Products.

Tools Used:

- Python, Pandas, Numpy, Matplotlib, Scikit-learn, Pytorch, Keras, Tensorflow, NLTK, Spacy, SQL, Amazon Web Services, Docker, Lambda, ECR, Step function, Sage Maker, Power BI, Teradata, SAS, Google Cloud Platform, Tableau.

Project 4 - IT Industry (42 months)

• Problem Statement:

1. ?Develop a python ML model to assess customer responses (sentiment analysis) from web site, social media and Bot which can help to improve customer satisfaction.
2. Develop Python Time Series Machine Learning models to predict expected customer interactions for future days, weeks, months & years with 92% accuracy.
3. Develop Python Text classification model to classify customer emails into Inquiry, Request and Complain to prioritize response.

• ML-Models:

1. ?Text mining, TextBlob, n-grams, word cloud.
2. Multiple techniques used based on Business type – MA, ARIMA and SARIMA.
3. NLTK – tf-idf vectorizer, truncated SVD, stop word, sent_tokenize, word_tokenize.

• Solution:

1. ?Found exiting service issues with Bank's processes & systems and improve customer experience.
2. Reduced the cost of customer service resources by keeping right size of resources based on model prediction.
3. Improved customer experience by handling complains on priority within timeline and reduce compensation paid to customers due to service disruption.

Achievements:

- Introduced New processes of Predictive Modelling using Python Machine Learning in Teleperformance India across all the sites in May 2013 and then helped existing 29 accounts forecasting teams to understand the calculations & factors considered in Python Time Series Forecasting using Machine Learning. All the accounts started maintaining their own processes by January 2014.
- Streamlined eight new accounts & five existing accounts Analytics & Data Management within three years while working with Teleperformance - Implemented Best Forecasting and Analytics processes in India, China, Mexico, Columbia, Brazil, Vilnius and Jakarta sites between June 2015 and August 2016.
- New analytics techniques helped management to take on time decisions and helped increase the yearly company (only for one account) revenue from 12M USD to 30M USD within one & half years.
- Worked with Data engineers to collect accurate and required data by improving ETL processes in Tier 3 databases from source systems to produce full customer view.

Responsibilities:

- Introduce innovative ways of Forecasting using Predictive Modelling (Python Machine Learning) & performance measurement dashboards which can help team to Forecast (Expected customer interactions) with 92%+ accuracy for different businesses – IT support (Adobe & SanDisk), Telecom (AT&T), Finance (Gain Credit formerly known as Global Analytics, UK), Transportation (Uber USA & India).
- Responsible for providing long term Data Analytics solutions which are easy to maintain and can bring change in business.
- Streamline data pipelines, collect data from multiple sources, analyse structured and unstructured data to equip management with real time decision making.
- Develop cost optimization model and revenue prediction model for global businesses.
- Develop NLP solutions to process large emails and chat data for classification and insight generation.

- Responsible for data management and its integrity to produce accurate insights.
- Implementations of new data analytics tactics based on business requirements.
- Co-ordination with Business heads to discuss analytics findings and get further directions to implement those solutions.
- Share business findings with Operations and Service Delivery Leaders which helps them to drive business Performance.
- Monitor and Drive Data Analytics Strategy at Aggregate and Functional level to achieve ideal Efficiency.
- Analyze large datasets and produce actionable insights.
- Help management to find business opportunities & lacking areas and suggest how to improve them based on statistical data analysis.
- Communicate with global stakeholders to implement data science solutions which helps in reduce cost and improve revenue.
- Share the business opportunities and lacking areas with Management & operations periodically based on statistical analysis.
- Create high level dashboard to see overall business performance for management.
- Co-ordinate with different teams to implement state of art Data Analytics, Business Intelligence, Predictive Modeling and Data Science solutions.

Tools Used:

- Python, Pandas, Numpy, Matplotlib, Scikit-learn,, NLTK, Textblob, Gensim, SQL, Microsoft Power BI, Tableau, Google Cloud Platform, Advanced Excel.

Project 5 - IT Industry (7 months)

Client Supported: IndiGo Airlines (Parent company of Interglobe Technologies Pvt. Limited)

Responsibilities:

- Designing & automating data analysis reports & Dashboards for better performance tracking and consultation.
- Maintain yearly, monthly and weekly Forecast for Airlines to meet Key Performance Indicators and Service level goals.
- Analyse trend in Volumes, Productivity, and other metrics to drive Staffing changes and Productivity Metric Improvements.
- Maintain weekly & monthly forecast of expected customer interactions to have accurate resources.

Tools Used:

- SQL, Advance Excel & Power Point

Project 6 - IT Industry (16 months)

- **Problem Statement:** Build fully automated end to end AI powered Solution to approve Personal Loan instantly without manual intervention to improve customer experience and reduce cost.
- **Data asked from customer while he/she applies on the website** - Demographics, Bank Statement, Salary slip, Selfie, Image of Office Entrance, Image of valuable item.
- **ML-Tech** Extracting Raw data from SQL (including output given by processed PDFs using AWS-Textract, Images processed by Google Vision API, CIBIL Score), Data Mining, Data Cleansing, Data Wrangling, Feature Selection, Feature Engineering, Statistical hypothesis testing, Sampling for Imbalanced Data, Logistic Regression & Decision Tree, Evaluating Bias Variance, Error Analysis, back testing, Score New Data.
- **Solution:**

1. As soon as customer applies on the website structured data gets stored in SQL and images & PDFs gets stored on Cloud storage & shown on Admin Panel→ Ops guy needs to go to “Admin Panel” click on process

PDF and process Images to get specific info -> Once images and pdfs processed by Google Vision & AWS-Textract within 42 secs admin guy needs to click on "Access Risk" then API calls the risk model by sending input features to AI-platform where model is deployed and returns the decision to admin panel and website with Approve/Reject/Refer and it also triggers email to customer & Ops guy accordingly.

2. ML is processing 54% applications without any manual work and 46% applications referred to operations team to evaluate and approve from admin panel as all the info flows to admin panel (Demographics, output from PDFs and Images).

Project 7 - IT Industry (6 months)

- **Problem Statement:** Design Metrics with statistical quantification to measure the digitization progress for all the Banking Channels and Products – CASA, Credit Card, Loans, FD, UT, RIB, MB, Branch, ATM, CDM.
- **ML-Tech** SQL for structured data collection & Python to clean unstructured data from website, email, social media, chatbot - integrated tables stored in EDW and Final real time output visualized in Tableau.
- **Solution:** Integrated structured & unstructured data from different sources for all channels and Products and produced real time visualisation in Tableau to measure the metrics i.e # of Transactions, Amount of Transactions, # of Unique Customers, Service Level, Downtime & Availability of Channels by Product & Segmentation based on transitional behaviour.

Awards

- Certification in Machine Learning from Udemy.
- Stanford Machine Learning course by Andrew Ng from Coursera.