AAKASH A

Contact: +91 9003215234 | E-mail: akash731@gmail.com | LinkedIn Link | Kaggle

LEAD DATA SCIENTIST

Data Science and ML: Clustering, Classification, Regression, Optimization, Model Selection, Dimensionality Reduction, Pre-processing, NLP, Deep Learning | Statistics and Mathematics: Probability, Probability Distributions, Hypothesis Testing

Core Skills: Predictive Analytics, Quantitative Research, Financial Modelling, Statistical Analysis, Reporting & Data Visualization, Data Analysis, Prescriptive Analytics, Requirement Gathering, Data Processing, Team Coordination

- Success in creating profitable business across Pharmaceutical, E-commerce, Retail, Call Centre for multiple countries in United States, Europe, Middle East, APAC region using Data Science, Machine Learning Models and Data Analytics.
- Expert in **designing, developing, and executing high-impact analytics solutions** for large, complex, structured, as well as unstructured data sets to help clients make better fact-based decisions.
- Know-how of utilizing statistical/mathematical techniques and knowledge of multiple industries to **design new algorithms and strategies and improve existing ones.**

Key Initiatives and Work Highlights

- Predicted patients for Amyloidosis and Breast cancer with an estimated revenue of **$25M annual impact in 2023-2024.**
- Provided call centre customer recommendation (clustering and classification) for Japan, the United Kingdom, and the United States resulted in a **$23M annual impact.**
- Helped the organisation create a **$2M impact per year** through online product recommendation engine for **APAC region.**
- Achieved rating and review aggregation across 26+ nations on an ecommerce site with an estimated annual impact of **$18M**.

PROFESSIONAL EXPERIENCES

Lead Data Scientist, AstraZeneca, Bengaluru, Karnataka | June 2023 to Present

Hewlett Packard Inc (HPI), Bengaluru, Karnataka | March 2018 to June 2023
Data Scientist III, Mar 2018 to Aug 2019 → Data Scientist, Aug 2019 to Aug 2020 → Lead Data Scientist, Aug 2020 to Jun 2023

Data Scientist, Sports Mechanics India Private Limited, Chennai, India| May 2017 to January 2018

Programmer Analyst (Data Science), Cognizant Technology Solution, Chennai, India | June 2015 to April 2017

EDUCATION AND CREDENTIALS

Bachelor of Engineering (BE), Computer Science & Engineering (CSE), Anna University, R.M.K Engineering College, Chennai | **2015**

Technical Skills: Programming Languages: Python, C++ ◆ Cloud Computing: Amazon Web Services (AWS) ◆ Databases: SQL, NO-SQL Database ◆ Big Data: Pyspark◆ Visualization Tools: Tableau ◆ Python Packages: Pandas, NumPy, SciPy, Scikit Learn, LGBM, XGBoost, Keras, TensorFlow, MLLIB, NLTK, Spacy, Pytorch

Certifications:

- Advance Machine Learning National Research University Higher School (www.coursera.com)
- Mathematics for machine Learning Imperial College London (www.coursera.com)
- Machine Learning Specialization by University of Washington (www.coursera.com)
- Machine Learning and Data Science Stanford University (Andrew NG) - (www.coursera.com)
- Deep Learning with Python (Lazy Programmer)-(www.udemy.com)
- Data Science and Machine Learning with Python - Hands On – (www.udemy.com)
- AWS Certified Solution Architect – Associate Level

COMPETITIONS, PROJECTS, PUBLICATIONS AND AWARDS

Awards and Recognition:
Best Data Scientist for the year 2020 at HP
Best Data Scientist SPOT award for 2021 at HP
Best candidate award for HP way for 2022 at HP

Competitions:
10th position in COVID19 Global Forecasting Week 1 at Kaggle with 87% accuracy
Stood top 100 in IEEE Fraud Detection at Kaggle with 90% accuracy

Publications:
"Element-Capacity Difference Algorithm for Optimal Bin Packing "selected to be published in International Journal of services and operations management.

Academic Projects:
"Highgriva.com" Start-up (2013-2015):
Stock Market Scrapper
Custom Scheduler

PROJECT ANNEXURE

Astrazeneca, Bengaluru, Karnataka | June 2023 to Present

- **Eplontersen:** Led a team of two individuals in developing a classification model to predict amyloidosis, a rare disease, using claims data. Utilized LightGBM with a custom objective function and K-means to predict early detection and the next visit for these patients.
- **Breast Cancer Drug:** Led a team of four individuals in creating an image classification and drug discovery model to assist metastatic breast cancer patients with Her2 and Brca mutations. The goal is to identify care gaps and deliver appropriate treatment for the elimination of cancer.

Hewlett Packard Inc (HPI), Bengaluru, Karnataka | March 2018 to Present

- **Program Management for Data Engineering and Data Science:** Lead a team of 4 people for creating a stable data environment for US, EMEA and Japan to scale the Machine Learning modules for enhanced performance of the models and also create new use cases for the business.
- **Customer Buy and Value Prediction** for SMB Customers (Japan and United Kingdom Call Centre): Used KMeans clustering, Random Forest, LSTM, XGBoost, Hyperopt, SMOTE and K-Fold Cross Validation to provide predictions at customer level. The Program is generating an incremental $5M revenue quarter on quarter from almost 1 year.
- **Product Personalization** for Ecommerce (India, Australia, Singapore, Thailand, and Indonesia): Developed a project that recommended product category for each customer who visited the website through banners, icons and advertisement live on the site. Used Catboost, XGBoost, LGBM, Hyper opt and K-Fold Cross Validation. The Program is generating an additional $1.5 Million revenue per year from the start of 2020.
- **Rating and Reviews** on Ecommerce Site (26 Countries): Grouped similar products to syndicate ratings and reviews across similar products. Used NLP, NLTK Spacy and AWS translate. Compiled distributed sharing of ratings and reviews across 26 countries and products. Created a engine to measure benefits which required working with click stream data from Adobe.
- **Order Cancellation** Prediction and Delivery Dashboard (United States and United Kingdom): Assigned a task to predict customers who would possibly cancel orders due to delays in delivery and visualize live geographic data as descriptive data so that business can optimize warehouses to fasten up delivery. Used KMeans clustering and ensembled classification. The predication saved almost $3M during black Friday and cyber-Monday sale.
- **Inventory Forecasting** to solve stockout issues (United States, India): Used algorithms such as LGBM Gradient boosting mode, RNN, Hyper opt, SMOTE and K-Fold CV for inventory forecasting. Generated additional $2M revenue quarter on quarter during sale and $1M revenue quarter on quarter during non-sale period from 2018.
- **SKU Rationalization**: Tasked to predict the maximum life of an SKU in the market to reducing overstocking and optimize product portfolio. Used algorithms such as Gradient Bosting Decision tree, K-Fold Cross validation and Integer Optimization. Saved $4.5M by eliminating non-selling SKU in 2018 and 2019.

Sports Mechanics India Private Limited, Chennai, India| May 2017 to January 2018

- **Indian Premier League Auction Engine** (Chennai Super Kings and Mumbai Indians): The aim of the project was to develop a team builder to create a team based on opponent player buy strategy. Used algorithms such as clustering, cosine distance, k-nearest neighbour, random forest regression ensemble classification, IBM-CPLEX. *The application aided in the development of Chennai Super King's victory strategy in 2018.*
- **Custom Conversational Chat Bot** for Feed Push to social media: Assigned a task to develop Chatbot to push scores to social media subscribed user for live scores. Used NLTK for Language processing, Stanford POS Tagger, Custom Noun Match Algorithm IBM Watson, Text Classification. *Increased 62% in live score subscription for local leagues on social media platform.*
- **Match Scheduler and Pitch Selection**: Worked on a project to build a match schedule and select pitches for matches in the future. Conducted distribution analysis and clustering (hierarchical clustering and spatial clustering). *Developed an application that provided unbiased match scheduler for multiple tournament and pitches.*
- **Player Ranker**: Tasked to use machine learning and statistics to develop an unbiased player ranker. Used TOPSIS - Customised statistical Ranking model, Support vector Machine, K-Fold Cross validation, Flask and D3JS. *Successfully created a player ranker which was used for multiple sue cases.*

Cognizant Technology Solution, Chennai, India | June 2015 to April 2017

- **Geneia Healthcare**: Worked on a project, that recommend the best medicine combination and schedule for Geneia patients. Used AWS ML, Apache Kafka, Apache spark and DynamoDB. *As a result of the application, 53% of patient successfully recovered faster for acute problems.*

- **State Auto**: Used Apache Spark, Spark ML, K-Means, Logistic Regression (Multiple Solvers) to analyse insurance data Upsell and cross sell insurance customised to user needs. *Successfully achieved an 42% increase in average subscription in 4 quarters.*
- **Siemens**: Undertook the project to develop a battery log parser for multiple devices to parse logs and perform tasks. Algorithms and service used were Pandas, NumPy and Amazon S3. *Developed a Recommender System that suggested the processes that must be terminated for efficient performance of the device.*