Nitin Jain

E-mail: njainiitp@gmail.com ● Contact: +91-7042353270 ● DOB: 12th October, 1989

Summary

- 5 years of experience as Data Scientist, with core expertise in architecting, pipelining and building Machine Learning based solutions.
- Tools: R, Spark, Hive, Hadoop, Jupyter notebook, Excel, Power Point
- Languages: Python, Scala, Hive, SQL, C#| Database: MySQL, Cassandra, Elastic search
- Areas of interest: Data mining, Machine learning (Scikit-learn, Rlibrary), statistical modelling, Hypothesis testing, Text mining, distributed parallel processing for large scale machine learning problems using Spark.
- Machine learning: Linear/ Logistic Regression, GLM, Segmentation, clustering (K-means, Hierarchical), KNN, Bagging, Boosting, Random forest, Support vector machines, Neural networks, Decision Trees

Work Experience

- **Manager- Data Science** at ***Times Internet*** (Jan'17-present)
  - Currently leading the efforts to establish machine learning and data science best practice for News business
  - Have set-up the Data science infrastructure for processing big data at large scale and use it for decision making across organization
  - Setting up the basic infrastructure to report important KPIs for business like Retention, cohort analysis, Segmentation, Segment profiling
  - Content personalization: Personalization of the news feed based on the historical clickstream behavior of the user, personalize first time experience for new users, recommend top publications based on user profile. (Using k-means clustering, collaborative filtering and content based approach)
  - Predict customers who are likely to churn based on user's clickstream behavior / App usage and take actions to pro-actively retain them. (Using ML techniques like Random Forest, XGBoost).

- **Data Scientist** at ***MakemyTrip.com*** (May'15-Jan'17)

  - **Propensity to buy a product:**
    - Used supervised learning algorithms to identify a target set of population who exhibit a high likelihood to purchase hotel or holiday over a future time.
    - CRM team used this set of customers to target with relevant offers and hence optimize the campaign ROI.
    - Built user level recommendations and offers were designed according to user need/references identified
  - **Customer churn prediction**
    - The objective was to increase the repeat purchase behavior of the customers.
    - Used various techniques like Decision trees, XGBoost and RF to predict the customer churn accurately
    - Created single 360 degree of the customers and consolidated data across all touchpoints of the customers with MMT and used supervised learning approaches to predict the churn.
    - Did exploratory data analysis to present the insights on customer repeat behavior to the CRM team
  - **Customer Lifetime Value**
    - Conceptualized the approach to be used to calculate cltv in the ecommerce domain. Being a non-contractual setting, this was challenging.
    - Applied Survival modelling and distribution based approach like Pareto/NBD and combination of XGBoost and CART models to predict the Customer Lifetime value.
    - Presented model validation results to CXOs, insights like who is a typical high LTV MMT customer, what is his age, gender, marital status, city tier, device preference, source of acquisition, affluence level and also insights like what kind of customers have we acquired lately during discounting vs non discounting etc.
    - Prepared use cases and driving implementation/usage by product and marketing teams.
  - **Increasing App engagement and user retention**
    - Business problem: A significant proportion of marketing budget goes into app downloads but eventually over a time period these apps mostly get uninstalled.
    - Identifying the drivers for App uninstalls and predicting the customers who are likely to uninstall App in near future.
    - Based on results, marketing team acted on these factors to retain customers, product team created different app on boarding program for different segment of new customers.

- **Analyst** at ***Fractal Analytics Pvt. Ltd.*** (Mar'14-May'15)

  Improve overall spend and revenue on credit cards and enhance offers and propositions for a leading Scandinavian bank

  - **Portfolio analysis :**
    - To provide bank with an overview of the spend and payment behavior on the current portfolio
    - Performed root cause analysis for the interesting observations in the portfolio
    - Presented recommendations and insights based on observations from the analysis
  - **Predict likelihood to revolve and revolve amount**
    - Predicted the probability of customers to generate interest revenue for the bank by building multiple logistic regression models for different segments of customers
    - Predicted the amount of interest revenue that a customer will generate for the bank by building linear regression models

- **Segmenting the customers:**
  Used K-means clustering to find the homogenous segments of customers based on their card spend on different merchant groups to better target them for different campaigns
- **Market basket analysis:**
  To find different combinations of merchant groups that has high affinity of being transacted together by customers. For e.g. customers spending on airlines is likely to spend on hotels also. So, Airlines and hotels have high co-affinity.

Enhancing customer lifecycle management: portfolio segmentation and credit card cross-sell

- **Acquisition**:
  - Identified customers among those not having debit card to sell debit card to them.
  - Developed cross-sell model to identify customer who hold debit cards and can be approached for credit card.
- **Activation**: Analysis to gain insights on activation behaviour of customers and proposed industry best practices to promote early activation.
- **Ongoing usage**:
  - Opportunity index: Based on capacity, activity and engagement parameters with different weights for each
  - Propensity to increase usage: by Logistic regression model
  - Segmentation based on above 2 parameters for better targeting of customers
- **Attrition**: Identified drivers for attrition using CHAID and provided recommendations to adopt proactive retention measures to reduce attrition of profitable customers.

- **Scientist** at *ARDE, Pune* (Jul'12-Mar'14)

  - Reporting and Analytics: Analysis of system performance data using excel, generate insights, preparing reports and presentations.
  - Consultation monitoring of client's problem statement and analysis
  - Designing based on client's requirement and data analysis using tools such as Excel and others
  - Optimizing the product design with necessary improvements obtained as reviews and feedbacks after testing and finally delivering it to the client

Education

- B.Tech in Mechanical Engineering from Indian Institute of Technology (**IIT**), Patna (2008-2012). **CGPA-8.57/10**.
- Senior secondary from Milton Academy, Bilaspur, U.P. (2007) with 76%
- Higher Secondary from Whitehall Public School, Rampur, U.P. (2005) with 81%

Awards/Achievements:

- Have been a consistent top performer: Awarded top 5 star rating in the current and previous organizations during annual appraisals.
- A keen data science enthusiast who is passionate about solving business problems through data, has participated in many online data science hackathons.
- Was among the top 5 students during the under graduation at IIT, was awarded merit scholarship all 4 years.