**Jay V**

**Data Scientist**

**1prtejay.ds@gmail.com**

# PROFESSIONAL EXPERIENCE

* Around 8 Years of Experience in developing different Statistical Machine Learning, Text Analytics, Data Mining solutions across various business functions: providing BI, Insights and Reporting framework to optimize business outcomes through data analysis.
* Developed and deployed dashboards in Tableau and R Shiny to identify trends and opportunities, surface actionable insights, and help teams set goals, forecasts and prioritization of initiatives.
* Experience using multiple ETL tools in Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export such as Ab Initio, Alteryx and Informatica Power Center
* Hands on experience in optimizing the SQL Queries and database performance tuning in Oracle, SQL Server and Teradata databases
* Strong mathematical knowledge and hands on experience in implementing Machine Learning algorithms like K-Nearest Neighbors, Logistic Regression, Linear regression, Naïve Bayes, Support Vector Machines, Decision Trees, Random Forests, Gradient Boosted Decision Trees, Stacking Models
* Expertise in transforming business requirements into building models, designing algorithms, developing data mining and reporting solutions that scales across massive volume of unstructured data and structured.
* Proficient in Machine Learning techniques (Decision Trees, Linear/Logistic Regressors, Random Forest, SVM, Bayesian, XG Boost, K-Nearest Neighbors) and Statistical Modeling in Forecasting/ Predictive Analytics, Segmentation methodologies, Regression based models, Hypothesis testing, Factor analysis/ PCA, Ensembles.
* Experience in designing visualizations using Tableau and PowerBI software and Storyline on web and desktop platforms, publishing and presenting dashboards.
* Hands on experience in implementing LDA, Naive Bayes and skilled in Decision Trees, Random Forests, Linear and Logistic Regression, SVM, Clustering, neural networks and good knowledge on Recommender Systems.
* Highly skilled in using visualization tools like Tableau, Matplotlib for creating dashboards.
* Extensive working experience with Python including Scikit-learn, Sci-py, Pandas and Numpy developing machine learning models, manipulating and handling data.
* Having good domain knowledge on B2B SaaS, Payment processing, Entertainment analytics, Healthcare, Retail.Well experienced in Normalization & De-Normalization techniques for optimum performance in relational and dimensional database environments.
* Regularly accessing JIRA tool and other internal issue trackers for the Project development.
* Integration Architect & Data Scientist experience in Analytics, Big Data, SOA, ETL and Cloud technologies.
* Worked and extracted data from various database sources like Oracle, SQL Server and Teradata.
* Experience in foundational machine learning models and concepts (Regression, boosting, GBM, NNs, HMMs, CRFs, MRFs, deep learning).
* Skilled in System Analysis, Dimensional Data Modeling, Database Design and implementing RDBMS specific features.
* Facilitated and helped translate complex quantitative methods into simplified solutions for users
* Knowledge of working with Proof of Concepts and gap analysis and gathered necessary data for analysis from different sources, prepared data for data exploration using data munging.
* Analyzed data using R, Perl, Hadoop and queried data using structured and unstructured databases
* Expertise in complete software development life cycle process that includes Design, Development, Testing and Implementation in Hadoop Eco System, Documentum sp2 suits of products and Java technologies.
* Experience developing data models and processing data through Big data frameworks like Hive, HDFS and Spark to access streaming data and implement data pipelines to process real-time suggestion and recommendations.

# TECHNICAL SKILLS

**Statistics:**

* **Exploratory Data Analysis:** Univariate/Multivariate Outlier detection, Missing value imputation, Histograms/Density estimation, EDA in Tableau
* **Supervised Learning:** Linear/Logistic Regression, Lasso, Ridge, Elastic Nets, Decision Trees, Ensemble Methods, Random Forests, Support Vector Machines, Gradient Boosting, Deep Neural Networks, Bayesian Learning
* **Unsupervised Learning:** Principal Component Analysis, Association Rules, Factor Analysis, K-Means, Hierarchical Clustering, Gaussian Mixture Models, Market Basket Analysis, Collaborative Filtering and Low Rank Matrix Factorization
* **Feature Selection:** Stepwise, Recursive Feature Elimination, Relative Importance, Filter Methods, Wrapper Methods and Embedded Methods
* **Statistical Tests:** T Test, Chi-Square tests, Stationarity tests, Auto Correlation tests, Normality tests, Residual diagnostics, Partial dependence plots and Anova
* **Sampling Methods:** Bootstrap sampling methods and Stratified sampling
* **Model Tuning/Selection:** Cross Validation, Walk Forward Estimation, AIC/BIC Criterions, Grid Search and Regularization
* **Time Series:** ARIMA, Holt winters, Exponential smoothing, Bayesian structural time series

**Machine Learning/ Deep Learning:**

* **R:** caret, glmnet, forecast, xgboost, rpart, survival, arules, sqldf, dplyr, nloptr, lpSolve, ggplot
* **Python:** pandas, numpy, scikit-learn scipy, stats models, ggplot2, tensorflow, Caffe, Theano, H2o, and Keras.
* **SAS:** Forecast server, SAS Procedures and Data Steps
* **Spark:** MLlib, GraphX
* **SQL:** Subqueries, joins, DDL/DML statements
* **Databases/ETL/Query:** Teradata, SQL Server, Postgres and Hadoop (MapReduce); SQL, Hive, Pig and Alteryx
* **Visualization:** Tableau, ggplot2 and RShiny
* **Prototyping:** RShiny, Tableau, Balsamiq and PowerPoint

# WORK EXPERIENCE

***Data Scientist Mar 2017–Present***

***Client: Macys, Atlanta, GA***

**Responsibilities:** Playing a key role in developing and maintaining statistical and machine learning models that mine, analyze and turn data into meaningful insights that is helping Macy’s grow their customer base, and revenue.

**Projects:**

**Market Mix Modeling:**

* Developed and refined complex marketing mix statistical models in a team environment and worked with diverse functional groups with over $100MM in annual marketing spend
* Responsible for all stages in the modeling process, from collecting, verifying, & cleaning data to visualizing model results, presenting results, and making client recommendations

**Customer Segmentation:**

* Developed 5 customer segments using unsupervised learning techniques like K Means and Gaussian mixture models. The clusters helped business simplify complex patterns to manageable set of 5 patterns that helped set strategic and tactical objectives pertaining to customer retention, acquisition, spend and loyalty.

**Sales Forecasting:**

* Implemented various advanced forecasting algorithms that were effective in detecting seasonality and trends in the patterns and thus helped in improving sale/demand forecast accuracy by 20-25% which helped business plan better with respect to budgeting and sales and operations planning
* Tuned model parameters (p, d, q for ARIMA) using walk forward validation techniques.

**Customer Churn Prediction:**

* Predicted the likelihood of customer attrition by developing classification models based on customer attributes like user demographics, historic clicks, user acquisition channels etc. The models deployed in production environment helped detect churn in advance and aided sales/marketing teams plan for various retention strategies in advance like tailored promotions and custom offers

**Cross Sell and Upsell Ad Opportunities:**

* Implemented market basket algorithms from transactional data, which helped identify ads clicked together frequently. Discovering frequent ad sets helped unearth Cross sell and Upselling opportunities and led to better pricing, bundling and promotion strategies for sales and marketing team

**Ad Click Propensity Modeling:**

* Developed machine learning models that predicted Ad click propensity of users based on attributes such as user demographics, historic click behavior and other related attributes. Predicting user propensity to click helped show and place relevant ads

**Customer Life Time Value Estimations:**

* Projected customer lifetime values based on historic customer usage and churn rates using survival models. Understanding customer lifetime values helped business to establish strategies to selectively attract customers who tend to be more profitable for Yahoo. It also helped business to establish appropriate marketing strategies based on customer values.

**Merchant Purchase Likelihood:**

* Developed a machine learning system that predicted purchase probability of a particular offer based on customer’s real time location data and past purchase behavior; these predictions are being used for mobile coupon pushes.

**Environment:**

Python, R, Machine learning algorithms, Hadoop/Spark, Tableau Desktop, Tableau server, Tableau Prep, SQL Server, T-SQL, NoSQL, Spark SQL, PySpark Mlib, SSIS, PowerBI

***Data Scientist Dec 2015 – Feb 2017***

***Client: Cox Communications, Atlanta, GA***

**Projects:**

**Price Elasticity Analysis:**

* Measured the price elasticity for products that experienced price cuts and promotions using regression methods; based on the elasticity, group on made selective and cautious price cuts for certain licensing categories.

**SKU Rationalization:**

* Developed algorithms for optimal set of Stock keeping units to be put in stores that maximized store sales, subject to business constraints; advised retailer to gauge demand transfer due to SKU deletion/addition to its assortment.

**Coupon Recommender System:**

* Developed a personalized coupon recommender system using recommender algorithms (collaborative filtering, low rank matrix factorization) that recommended best offers to a user based on similar user profiles. The recommendations enabled users to engage better and helped improving the overall user retention rates at Nordstrom

**Store Supply Chain Segmentation:**

* Clustered the supply chain of Nordstrom stores based on volume, volatility in demand and proximity to warehouses using Hierarchical clustering models and identified strategies for each of the clusters to better optimize the service level to stores

**Campaign Effectiveness Measurement:**

* Built Tableau dashboards that tracked the pre and post changes in customer behavior post campaign launch; the ROI measurements helped retailer to strategically extend the campaigns to other potential markets

**Customer Trial & Repeat Behavior Analysis:**

* Designed and deployed real time Tableau dashboards that identified items which are most/least liked by the customers using key performance metrics that aided retailer towards better customer centric assortments. It also aided retailer towards strategies pertaining to better ad placement, bundling and assortments

**Environment:**

Python, R, Machine learning algorithms, Hadoop/Spark, Tableau Desktop, Tableau server, Tableau Prep, SQL Server, T-SQL, NoSQL, Spark SQL, PySpark Mlib, SSIS, PowerBI

***Data Scientist Feb 2013 – Oct 2015***

***Client: Bank of America, Hyderabad, India***

**Responsibilities:** Played key role in developing and deploying Dodd-Frank Act Stress Test models across several bank portfolios. Provided architectural leadership on several high priority initiatives including account prioritization, account prospecting, and opportunity scoring. Drove the creation of comprehensive datasets encompassing user profiles and behaviors and incorporating a wide variety of signals and data types.

**Projects:**

**Forecasting Loan Balance:**

* Forecasted bank-wide loan balances under normal and stressed macroeconomic scenarios using R. Performed variable reduction using the stepwise, lasso, and elastic net algorithms and tuned the models for accuracy using cross validation and grid search techniques.

**Top down Models (Commercial Real Estate):**

* Automated the scraping and cleaning of data from various data sources in R and Python. Developed Banks' loss forecasting process using relevant forecasting and regression algorithms in R.
* The projected losses under stress conditions helped bank reserve enough funds per DFAST policies

**Default Loan Prediction:**

* Built classification models using several features related to customer demographics, macroeconomic dynamics, historic loan payment behavior, type and size of loans, credit scores and loan to value ratios and with accuracy of 95% accuracy the model predicted the likelihood of default under various stressed conditions.

**Credit Risk Scorecards:**

* Built credit risk scorecards and marketing response models using SQL and SAS. Evangelized the complex technical analysis into easily digestible reports for top executives in the bank.
* Developed several interactive dashboards in Tableau to visualize d2 Terabytes of credit data by designing a scalable data cube structure.

**Environment:**

Python, R, Machine learning algorithms, Hadoop/Spark, Tableau Desktop, Tableau server, Tableau Prep, SQL Server, T-SQL, NoSQL, Spark SQL, PySpark Mlib, SSIS, PowerBI

***Data Analyst Jun 2011 – Jan 2013***

***Client: CTE Global Solutions, Hyderabad, India***

**Project:** The company is a leading participant in vendor financing, factoring, equipment and transportation financing, Small Business Administration loans, and asset-based lending. The Goal of this project was to do analysis of delinquencies and rejection in loan process.  
  
**Responsibilities:**

* Created views in Tableau Desktop that were published to internal team for review and further data analysis and customization using filters and actions.
* Created Heat Map showing current customers by color that were broken into regions allowing business user to understand where we have most users vs. least users using Tableau.
* Projected and forecasted future growth in terms of number of customers in various classes by developing Area Maps to show details on which states were connected the most and publishing it on Tableau Server.
* Converted charts into Crosstabs for further underlying data analysis in MS Excel.
* Created Bullet graphs to determine profit generation by using measures and dimensions data from Oracle, SQL Server and excel.
* Blended data from multiple databases into one report by selecting primary key from each database for data validation.
* Combined views and reports into interactive dashboards in Tableau Desktop that were presented to Business Users, Program Managers, and End Users.
* Developed story telling dashboards in Tableau Desktop and published them on to Tableau Server which allowed end users to understand the data on the fly with the usage of quick filters for on demand needed information.
* Tested dashboards to ensure data was matching as per the business requirements and if there were any changes in underlying data.
* Rewrote various business process and tested result in MS Excel using various functions and sub query with not exists.
* Involved in updating functional requirement document after development and created documentation for deployment team.
* Data quality check on variable level including missing values, unique values, frequency tables.
* Obtained the data from variety of sources such as Database, CSV, flat files etc.
* Wrote complex join SQL queries to extract, load data.

**Environment:**

MS SQL Server, PL-SQL, TOAD for Data Analyst, MS Project, MS Visio, Tableau