**RIMEETH K**

**DATA SCIENTIST**

**rimeeth1483@gmail.com**

**MOBILE: (931)-354-2574**

**SUMMARY:**

Passionate Data Scientist having strong background in statistics with 8+ years of experience, seeking to solve the boatnecks of the business and increase the productivity by implementing the Machine Learning models. Skilled in predictive modelling, data processing and data mining algorithms.

**PROFESSIONAL SUMMARY:**

* Around 8+ years of experience in IT field and 3+ years of experience as **Data scientist** with strong technical expertise, business experience, and communication skills to drive high-impact business outcomes through data-driven innovations and decisions.
* Expertise in **Statistical analysis, Predictive modeling, Text mining, Supervised learning, Unsupervised Learning,** and **Reinforcement learning.**
* Proficient in Statistical methodologies such as **Hypothesis Testing, ANOVA, Monte Carlo Sampling** and **Time Series Analysis.**
* Strong mathematical background in **Linear algebra, Probability, Statistics, Differentiation and Integration.**
* Expertise in handling with the various sets of data sources such as **semi structured, unstructured, time series,** and **spatial data.**
* Experience in balancing the datasets by using Resampling methods such as **Oversampling** and **Under sampling** techniques.
* Proficient in data visualization tools such as **Tableau and Python Matplotlib(Iplotly**) to create visually powerful and actionable interactive reports and dashboards.
* Excellent knowledge in numerical and scientific libraries such as **SciPy** and **NumPy.**
* Excellent understanding of Analytics concepts and Supervised Machine Learning algorithms like **Logistic Regression, Linear regression, K-Nearest Neighbors, Naïve Bayes, Support Vector Machines, Decision Trees and Ensemble models: Random Forests, Gradient Boosted Decision Trees, Stacking Models**.
* Excellent understanding of Analytics concepts and Unsupervised Machine Learning algorithms like **K-Means, Density Based Clustering (DBSCAN), Hierarchical Clustering** (**Agglomerative, Divisive**) and good knowledge on **Recommender Systems.**
* Proficient in implementing **Dimensionality Reduction Techniques** like **Principal Component Analysis, t-Stochastics Neighborhood Embedding (t-SNE),** and **Linear Discriminant Analysis(LDA).**
* Experience with Deep Learning net families, such as **Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks** and **GANs.**
* Experience in building models with **TensorFlow** andtop-level frameworks such as **Keras, Theano and Pytorch.**
* Actively involved in all phases of data science project life cycle including **Data Extraction, Data Cleaning, Data Visualization** and **building Models**.
* Experience in working with languages **Python** and **R.**
* Strong knowledge in all phases of the **SDLC (Software Development Life Cycle)** from analysis, design, development, testing, implementation and maintenance.
* Experienced in Agile methodology and SCRUM process.
* Experience in Extract, Transfer and Load process using ETL tools like **Data Stage**, **Data** **Integrator** and **SSIS** for Data migration and Data Warehousing projects.
* Experienced in Data Integration Validation and Data Quality controls for ETL process and Data Warehousing using **MS Visual Studio**, **SSAS, SSIS** and **SSRS**.
* Well experienced in Normalization, De-Normalization and Standardization techniques for optimal performance in relational and dimensional database environments.
* Experience in working at **Low Latency application** with performance tuning on the **Regularization.**
* Extensive working experience with python including **Scikit-Learn**.
* Experience in working on both **windows, Linux** platforms.
* Experience in using **GIT Version Control System.**

**PROFESSIONAL EXPERIENCE:**

|  |  |
| --- | --- |
| **Client: Sprint, Kansas City, KS** | **Sep 2017 – Till date** |
| **Role: Data Scientist** |  |

**Description:** Sprint Corporation is an American [telecommunications](https://en.wikipedia.org/wiki/Telecommunications) [company](https://en.wikipedia.org/wiki/Company) that provides [wireless services](https://en.wikipedia.org/wiki/Wireless_telecommunications) and is an [internet service provider](https://en.wikipedia.org/wiki/Internet_service_provider).  It is the fourth-largest [mobile network operator](https://en.wikipedia.org/wiki/Mobile_network_operator) in the United States and serves 54 million customers as of October 2017. headquartered in Kansas City, Kansas.

**Responsibilities:**

* Responsible for Data Cleaning, features scaling, features engineering by using **NumPy** and **Pandas** in Python.
* Performed Exploratory Data analysis **(EDA)** to maximize insight in to the dataset, detect the outliners and extract important variables by graphically and Numerically.
* Implemented algorithms such as **Principal Component Analysis** **(PCA)** and **t-Stochastics Neighborhood Embedding (t-SNE)** for dimensionality reduction and normalize the large datasets.
* developed various **Clustering algorithms** for market segmentation to analyze the customer behavior patterns.
* Used **Pandas, NumPy, seaborn, SciPy, Matplotlib, Seaborn, Scikit-learn**, **NLTK in Python** at various stages for developing machine learning model.
* Implemented machine learning algorithms, **Random forest** and **Support vector machines** to predict the Customer churn and Customer interface.
* Used **cross-validation techniques** to avoid the overfitting of the model to make sure the predictions are accurate and measured the performance using **Confusion matrix** and **Classification report.**
* Improved accuracy using Ensemble methods of the training model with different **Bagging** and **Boosting** methods.
* Performed Text analytics on unstructured email data using **Natural language processing tool kit(NLTK).**
* Involved in various pre-processing phases of text data like **Tokenizing**, **Stemming, Lemmatization** and converting the raw text data to structured data
* Performed feature engineering, performed NLP by using some techniques like **Word2Vec, BOW** (Bag of Words)**, tf-idf, Avg-Word2Vec**, **if-idf**, Weighted **Word2Vec.**
* Performed Sentimental analysis(**NLP**) on the email feedback of the customers to determine the emotional tone behind the series of words and gain the express of the attitudes and emotions by Long-Short Term Memory (**LSTM)** cells in **Recurrent Neural Networks(RNN).**
* Implemented **LSTM layer network** of moderate depth to gain the information in the sequence with help of **Tensor Flow.**
* Created **distributed environment** of **Tensor Flow** across **multiple devices** (CPUs and GPUs) and run them in parallel.
* Used PySpark Machine learning library to build and evaluate different models.
* Used Tableau to convey the results by using dashboards to communicate with team members and with other data science teams, marketing and engineering teams.
* Communicated the results with operations team for taking best decisions.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, SQL, Linux, Git, Microsoft Excel, PySpark-ML, Random Forests, SVM, t-SNE, PCA, Tensor Flow, K-Means, Natural Language Tool Kit, LSTM - RNN

|  |  |
| --- | --- |
| **Client: Fidelity Investments, Durham, NC** | **Sep 2016- Sep 2017** |
| **Role: Data Scientist** |  |

**Description:**  is a [multinational](https://en.wikipedia.org/wiki/Multinational_corporation) [financial services](https://en.wikipedia.org/wiki/Financial_services) corporation. It is the fourth largest asset manager with $2.4 trillion in [assets under management](https://en.wikipedia.org/wiki/Assets_under_management) as of December 2017. Fidelity Investments operates a [brokerage firm](https://en.wikipedia.org/wiki/Brokerage_firm), manages a large family of [mutual funds](https://en.wikipedia.org/wiki/Mutual_fund), provides fund distribution and investment advice, retirement services, [wealth management](https://en.wikipedia.org/wiki/Wealth_management), securities execution and clearance, and [life insurance](https://en.wikipedia.org/wiki/Life_insurance).

**Responsibilities:**

* Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database.
* Responsible for Data Cleaning, features scaling, features engineering by using **NumPy** and **Pandas** in Python.
* Performed Exploratory Data analysis **(EDA)** to maximize insight in to the dataset, detect the outliners and extract important variables by graphically and Numerically.
* Applied resampling methods like Synthetic Minority Over Sampling Technique (**SMOTE)** to balance the classes in large data sets.
* Implemented algorithms such as **Principal Component Analysis** **(PCA)** and **t-Stochastics Neighborhood Embedding (t-SNE)** for dimensionality reduction and normalize the large datasets.
* Implemented classification algorithms such as **Logistic Regression**, **K-NN neighbors** and **Random Forests** to predict the Customer credit history and Payments activity on historical data to get predicted label whether the customers are eligible for Credit card offers and report the marketing team.
* Built machine learning models and constructed multilayer perceptions for Deep Neural Networks **(DNN)** to identify fraudulent applications for loan pre-approvals and to identify fraudulent credit card transactions using the history of customer transactions and compared the results.
* Perform model tuning to find the best hyperparameter fit for the algorithm to achieve better results
* Ensemble methods were used to increase the accuracy of the training model with different Bagging and Boosting methods
* Performed data visualization and Designed dashboards with Tableau, and generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.
* measured the performance using **Confusion matrix** and **Classification report.**
* Evaluated models using **Cross Validation, Log loss function, ROC Curves** and **AUC** for feature selection.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, SQL, Linux, Git, Microsoft Excel, Random Forests, SVM, t-SNE, PCA, Tensor Flow, Python, DNN, K-NN

|  |  |
| --- | --- |
| **Client: Flipkart, Bangalore, India,Karnataka** | **April 2014- June 2016** |
| **Role: Data Scientist** |  |

**Description: Flipkart Pvt Ltd.** is an Indian [electronic commerce](https://en.wikipedia.org/wiki/Electronic_commerce) company based in [Bengaluru](https://en.wikipedia.org/wiki/Bengaluru), India. the company initially focused on book sales, before expanding into other product categories such as [consumer electronics](https://en.wikipedia.org/wiki/Consumer_electronics), fashion, and lifestyle products. Flipkart held a 39.5% market share of India's e-commerce industry.

**Responsibilities:**

* Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database.
* Gathered, analyzed, and translated business requirements, communicated with other departments to collected client business requirements and access available data.
* Responsible for Data Cleaning, features scaling, features engineering by using **NumPy** and **Pandas** in Python.
* Conducted Exploratory Data Analysis using Python **Matplotlib** and **Seaborn** to identify underlying patterns and correlation between features.
* Used information value, principal components analysis, and Chi square feature selection techniques to identify.
* Applied resampling methods like Synthetic Minority Over Sampling Technique (**SMOTE)** to balance the classes in large data sets.
* Designed and implemented customized **Linear regression model** to predict the sales utilizing diverse sources of data to predict demand, risk and price elasticity.
* Experimented with multiple classification algorithms, such as **Logistic Regression**, **Support Vector Machine (SVM)**, **Random Forest**, **AdA boost and Gradient boosting** using Python Scikit-Learn and evaluated the performance on customer discount optimization on millions of customers.
* Used **F-Score, AUC/ROC**, **Confusion Matrix** and **RMSE** to evaluate different model **performance**
* Performed data visualization and Designed dashboards with **Tableau**, and generated complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders.
* Used **Keras** for implementation and trained using cyclic learning rate schedule.
* Overfitting issues was resolved by batch norm, dropout helped to overcome the issue.
* Conducted in-depth analysis and predictive modelling to uncover hidden opportunities; communicate insights to the product, sales and marketing teams.
* Built models using **Python** and **Pyspark** to predict the probability of attendance for various campaigns and events.

**Environment:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, SQL, Linux, Git, Microsoft Excel, PySpark-ML, Random Forests, SVM, Tensor Flow, Keras

|  |  |
| --- | --- |
| **Client: Innova Solutions Pvt Ltd, Hyderabad, Telangana** | **na Apr 2011- Mar 2014** |
| **Role: SR. Tableau/SSIS Developer** |  |

**Description:** Innova Solutions is passionately client‑focused with decades of deep experience solving   
real‑world problems using advanced technology and approaches. Over more than 1000 professionals work in the industry.

**Responsibilities:**

* The goal of the project is to extract data from various heterogeneous sources and load that data into data warehouse using SSIS and generate reports using Tableau.
* Involved in complete Software Development Life Cycle **(SDLC)** process by analyzing business requirements and understanding the functional work flow of information from source systems to destination systems.
* Involved in complete SSIS life cycle in creating **SSIS** packages, building, deploying and executing the packages in both the environments (Development and Production).
* Created SSIS packages for transferring data from various data sources like SAP R3 system, Oracle, MS Access, Excel, \*. Txt file, \*.CSV files.
* Used various Transformations in **SSIS Dataflow**, **Control Flow** using for loop Containers and Fuzzy Lookups etc.
* Created Complex **ETL** Packages using SSIS to extract data from staging tables to partitioned tables with incremental load.
* Designed high level **ETL architecture** for overall data transfer from the source server to the Enterprise Services Warehouse which encompasses server name, database name, accounts, tables and direction of data flow, data dictionary and Metadata using SSIS.
* Worked on different levels of ETL Loads like Extract Data from Source System to Staging and then from Staging to load data in Database Tables.
* Involved in developing and deploying data transfers packages to the appropriate destinations using SSIS.
* Used various SSIS tasks such as Conditional Split, Derived Column, lookup which were used for Data Scrubbing, data validation checks during Staging, before loading the data into the Data warehouse.
* Used SSIS and **T-SQL** stored procedures to transfer data from OLTP databases to staging area and finally transfer into data warehouse.
* Involved in writing complex **SQL Queries, Stored Procedures, Triggers, Views, Cursors, Joins,** **Constraints, DDL, DML** and User Defined Functions to implement the business logic and created clustered and non-clustered indexes.
* Involved designing technology roadmap for Tableau: product installation and implementation development of insightful dashboards; serving as SME point of contact for Tableau and delivery of solutions adhering to **BI** industry best practices.
* Created **Tableau dashboards** using stack bars, bar graphs, scattered plots, geographical maps, Gantt charts etc. using show me functionality.
* Created interactive data visualizations in Tableau, using relational and aggregate data sources.
* Created **complex dashboard** using parameters, sets, groups and calculations and drill down and drill up in worksheets and customization using filters and actions.
* Developed **story telling dashboard**s 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.
* Tableau administration and development roles and helping the client with the issues of data connections, extraction and maintenance.

|  |  |
| --- | --- |
| **Client: Genpact, Hyderabad, Telangana** | **Sep 2009- APR 2011** |
| **Role: SQL Developer** |  |

**Description:** Genpact began in 1997 as a unit within [General Electric](https://en.wikipedia.org/wiki/General_Electric). Its charter was to provide business process services to GE's businesses. During the eight years that followed, Genpact began to manage a wide range of processes across GE's financial services and manufacturing businesses.

**Responsibilities:**

* Involved in complete Software Development Life Cycle **(SDLC)** process by analyzing business requirements and understanding the functional work flow of information from source systems to destination systems.
* Developed database schemas, tables and dictionaries.
* Developed Stored Procedures, functions and database triggers and maintained referential integrity and implement complex business logic.
* Created advanced select statement to retrieve data from various tables for business analysis.
* Created and maintained database objects, created complex Stored Procedures, Tables, Views & SQL Join statements for various applications.
* Created SSIS packages for transferring data from various data sources like SAP R3 system, Oracle, MS Access, Excel, \*. Txt file, \*.CSV files.
* Developed database monitoring and tuning strategies. Tuned Complex SQL Queries.
* using complex T-SQL statements, and implemented various types of Constraints and Triggers for Data Consistency.
* Performed tuning and optimization of queries for reports that take longer execution time using SQL Profiler, index tuning wizard and SQL Query Analyzer.
* Created User Defined functions and used it for creation of stored procedures.
* Worked on Database Backup, Recovery and Disaster Recovery Procedures.
* Used SQL Profiler to trace slow running queries and fixed them accordingly.
* Tested databases and performed bug fixes.
* Provided authentication and authorizations for Database.

**EDUCATION:**

Master of science in engineering management Marshall University

Bachelor of Technology in Mechanical Engineering Jawaharlal Nehru Technological University(JNTU)

**TOOLS AND TECHNOLOGIES:**

|  |  |
| --- | --- |
| **Languages** | C, C++, R, Python |
| **Databases** | Oracle 11g/10g, SQL Server, MS-Access, SSIS |
| **Mathematical** | Matrix operations, Differentiation, Integration, Probability, Statistics,  Linear Algebra, Geometry. |
| **Machine Learning Algorithms** | Logistic Regression, Linear Regression, Support Vector Machines, Decision  Trees, K-Nearest Neighbors, Random Forests, Gradient Boost Decision Trees, Stacking Classifiers, Cascading Models, Naïve Bayes, K-Means Clustering, Hierarchical Clustering and Density Based Clustering. |
| **Machine Learning Techniques** | Principal Component Analysis, Truncated SVD, Data Standardization,  L1 and L2 Regularization, Loss Minimization, Hyper Parameter Tuning, Performance Measurement of Models, Featurization and Feature Engineering, Content Based and Collaborative Based Filtering, Matrix Factorization, Model Calibration, productionizing Models, A/B Testing, Point and Interval Estimation, Hypothesis Testing, Cross Validations, Decision Surface Analysis, Retraining, Models periodically, t-stochastic neighborhood embedding. |
| **Deep Learning** | Artificial Neural Networks, Convolutional Neural Networks, Multi-Layer  perceptron’s, Recurrent Neural Networks, LSTM, GRU, SoftMax Classifier,  Back Propagation, Chain Rule, Choosing Activation Functions, Drop out,  Optimization Algorithms, Vanishing and Exploding Gradient, Striding, Padding, Optimized weight Initializations, Gradient Monitoring and Clipping, Batch Normalizations, Max Pooling. |
| **SQL Tools** | SQL Server Tools SQL Server Management Studio (SSMS), Erwin Data Modeler, SAS, Data Warehouse Tools MS SQL Server 2005/2008/2012/2014/2016 Integration Services Business Intelligence Tools SQL Server 2005/2008/2012/2014 Business Intelligence Development Studio Reporting Tools SQL Server 2005/2008/2012/2014 Reporting Services |
| **Visualization tools** | Tableau, Plotly |
| **Operating systems** | Linux, Windows |
| **Methodologies** | Agile waterfall |