

SUBRAMANIAN SHANMUGAVEL

201 S, 4th Street, Apt 405

San Jose, CA-95126

Subramaniankalyan@gmail.com • (812)369-8629.

Professional Summary

Over 4 years of Strong experience in **Data Analysis** with newly acquired skills (**Analytical SQL, Machine Learning, NLP/ Text Mining, Data Analytics, Data Visualization and BI**), an insatiable intellectual curiosity, and the ability to mine hidden business insights located within large sets of structured, semi-structured and unstructured data.

Strong Experience writing **SQL queries** for **data analysis and data evaluation**.

Hands on experience in implementing advanced analytical functions **Lead/Lag, Over, Partition By, Dense Rank, N-Tile, Row Number, Windowing Aggregate Function, Inverse Percentile, First/Last Functions**.

Proficient in Statistical Modeling and Machine Learning techniques (**Linear, Logistics, Decision Trees, Random Forest, SVM, K-Nearest Neighbors, Bayesian, XG Boost**) in Forecasting/ Predictive Analytics.

Demonstrated ability to work actively in different phases of **Software Development Life Cycle (SDLC)** using **Agile and Waterfall methodologies**.

Experience in report writing using **SQL Server Reporting Services (SSRS)** and creating various types of reports like **drill down, Parameterized, Cascading, Conditional, Table, Matrix, Chart and Sub Reports**.

Extensive experience in Analytics, developing different **Statistical Machine Learning, Data Mining** solutions to various business problems and generating data visualizations using **R, Python and Tableau**.

Experience in data driven Statistical analysis like Sampling, finding the Data distribution, **Hypothesis testing, Correlation among the variable, Outlier Detection, Analysis of Variance and Probability theory**

Strong understanding of **Data Modeling and Data mining**.

Highly skilled in using visualization tools like **Tableau, Matplotlib and ggplot2** for creating dashboards.

Excellent knowledge of **Hadoop Ecosystem** and Big Data tools as **Pig, Hive & Spark**.

Compile and validate data; reinforce and maintain compliance with corporate standards.

Handled importing data from various data sources, performed transformations using **Hive, Map Reduce, and loaded data into HDFS**.

Created PL/SQL packages, **Procedures, Functions, Triggers**.

Experience in developing **UML use case diagrams, Work Flow Diagrams** using tools like **Rational Rose and MS Visio**

Strong communication skills, **Quick learner and Team player**.

Skills & Abilities

- **Languages: Proficient:** SQL, Python, R, C++ **Intermediate:** SAS, JAVA
- **Databases:** Oracle (PL/SQL), MySQL, Microsoft SQL Server, PostgreSQL, Apache Hadoop
- **Tools:** Spark, Pig, Hive, MS Excel, SSIS, SSDT, SSMS, Tableau, Google Analytics, AWS
- **Packages:** Scikit-learn, pandas, Tensor flow, Matplotlib, Seaborne, ggplot2, plotly, Keras.
- **Math & Stats:** Logistic Regression, Linear Regression, Time Series Analysis, Hypothesis Testing, Analysis of Variance (ANOVA), Chi-Square Test, Z-Test and T-Test, Predictive Modeling, Numerical Analysis.

Certifications

- **Python & R for Business Analysis & Visualization**
- **Hands -On Tableau Training for Data Science**
- **The Ultimate Hands-On Big Data (Hadoop, Hive, Spark, Cassandra, Pig)**
- **Statistics for Data Science and Business Analysis**
- **A-Z Machine Learning for Data Science**

Experience

Data Analyst | UITS, IN | Bloomington | Jun 2017-Dec 2017

- **Instructure** - Aggregates **students data** enrolled across **Institutions**, providing new understanding of the student experience and **insights** into making it more successful. Insights into **student performance** to improve retention and graduation rates.
- My job is to capture and **organize** real-time data for the optimal use and to develop a strong data-based learning **analytics solution** to enable faculty and administrators to proactively identify, understand and act on student behavior and performance, empowering the right students at the right time to achieve greater **success**.
- Created **Visualizations, Dashboards, and reports** utilizing **Tableau** and interpret trends or patterns that tell a compelling story.
- **Acquire** and **Integrate** data from multiple sources using **Denodo** and by formulating complex **SQL Queries**
- Write Business Requirements documentation in **Word and Excel**, maintained in SharePoint.
- Work with **Data Architect** on Dimensional Model with both **Star** and **Snowflake Schemas** utilized.
- Performed **Data Analysis** on both source data and target data after transfer to **Data Warehouse**.
- Created **Tables, Views, Sequences, Indexes, Constraints** and generated SQL scripts for implementing physical data model.
- Environment: **SQL, Python, Tableau, Hive, Spark, Denodo, Redshift, Matplotlib, Pandas**

Data Analyst | Dynamic Solutions | India | Sep 2013 – Apr 2016

Applied Data Finance

- **Data Extraction and analysis** were done using TransUnion CIBIL credit reports where CIBIL is one of the largest collection of consumers Data globally.
- **Analysis** were done before giving approval for the loan, related to factors like Credit Score, Homeownership, Purpose of the loan, Term length of the loan, loan amount, income, Employee or not, Previous bankruptcy fraud.
- Implemented advanced analytical functions **Lead/Lag, Over, Partition By, Dense Rank, N-Tile, Row Number, Windowing Aggregate Function, Inverse Percentile, First/Last Functions**.
- Data Handling, Maintenance and Analysis were done using **Python and Tableau**.
- Followed working knowledge of **waterfall and agile methodologies** to be a proven factor for maximizing success.
- Ability to multi-task, prioritize and manage multiple simultaneous projects under tight and conflicting deadlines.
- Involved with Data Analysis Primarily Identifying **Data Sets, Source Data, Source Meta Data, Data Formats and Data Definitions**.
- **Predicted** the risk of the customers, the responsiveness of the targeted population and predicted the appropriate customers for the business from that information.
- Environment: **SQL, Python, Tableau, Sckit-learn, Matplotlib, Pandas, Seaborn, SSIS, SQL Server, Statistics, AWS, Hadoop**

Research Projects and Publications

Student Connect

Mar 2016 – Sep 2016

- The aim of the project was to provide outlay for the **infrastructure of a system** which would help university students to communicate.
- We devised a **system** and elaborate **infrastructure** layout for resolving most of the problems which will group students based on various interests.
- Acquired data from various departments and student bodies regarding the problems students face by conducting **interviews and surveys**.
- Worked with a varied range of modeling languages to create **structural, behavioral, and interaction level diagrams**.

- Coordinated with **ETL** team and **QA** team to improve the quality of the **model** and **reports**(analytics)
- Worked with a diverse team to gather requirements from **Office of International Services**.

An Adaptive approach of Tamil Character recognition using Deep Learning (Indic- Computing)

Jul 2015 -Dec2016

- Conducted deep learning research to develop a Tamil character recognition system using Convolutional neural network and published at © Springer international publishing Switzerland and presented at CSI-2015.
- Classify characters in Tamil, a south Indian language, using Convolutional Neural Networks (ConvNets).
- Applied Regression, Clustering, and Classification techniques to understand correlation and performance measures.

Improving Restaurants on Yelp by Sentiment Analysis of Reviews

Jan 2015 – Jul 2015

- Answered several business questions by using sentiment analysis and various scikit-learn methods on the Yelp restaurant reviews and attributes.
- Explored the reviewer's perspectives through extracted phrases and how restaurants can make use of it to improve their businesses.
- Explained the correlation between the average polarity scores obtained and the star ratings
- Delivered the classification of reviews and how these phrases extracted were visualized.

MEDLINE

Aug 2017 – Dec 2017

- An approach that retrieved and organized MEDLINE citations into different topical groups and prioritized important citations in each group.
- Create a Clustered search on PubMed corpus of medical/bio-medical journals using HPC at IU
- Effectively integrated Text clustering, Text summarization, and Text ranking and organized MEDLINE retrieval results into different topical groups.
- Unsupervised learning – Document clustering, indexing and search, word embeddings, feature extraction.

IUB-Microsoft Funded Project

Jan 2017 – Dec 2017

- To deliver a web-deployable Handwritten Text Recognition system by means of which repositories of scientific data can render handwritten text on specimen labels and ledgers into digitized information.
- Develop a model to correctly predict digital encoding for each handwritten character in the catalog and specimen label images, with maximum accuracy feasible Technology like Computer Vision & Image Processing.

Education

MASTER OF SCIENCE | Indiana University Bloomington, IN, USA

- **Major:** Data Science

Master of Engineering | Anna University, TN, India

- **Major:** Computer & Communication

Bachelor of Technology | Anna University, TN, India

- **Major:** Information Technology