

SURESH KUMAR

DATA SCIENTIST

sureshynp9@gmail.com

Contact: 91 -9880420554



An IT professional with over 4.4 years of experience as a Data Scientist and Bigdata Support Engineer. Data scientist models build using Machine Learning, Statistical Modeling, Data Mining, Natural Language Processing and Data Visualization and using R, Python, Tableau.

Professional Summary

- Good working knowledge in supervised, unsupervised Machine learning algorithms like Linear Regression, Logistic Regression, K Nearest Neighbours, Support vector machines, Naives Bayes classification, Decision Trees, Random Forest, Clustering techniques, anomaly detection, Time series analysis.
- Collaboration and interaction with other teams: Working closely with data management colleagues to guide them on structuring relevant data to facilitate data exploration and fast prototyping, working closely with business insight managers to ensure business relevance of analytics processes and products, working with data engineers on incorporating data from external sources which would complement internal data.
- Proactive participation in products roadmap discussions, data science initiatives and the optimal approach to apply the underlying algorithms.
- Collaborated with data engineers and operation team to implement the ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
- Participate in business reviews to provide inputs and identify commercial opportunities for Data Science.
- Utilize in-depth knowledge of functional and technical experience in Text Mining, Data Mining, Data Pre-processing, Data Scientist, and Tableau concepts and business skills to deliver solutions to customer.
- Ability to play a key role in the team and communicates across the team.
- Good working knowledge of text analytics, data mining, using NLP
- Well versed writing production ready code in Python and SQL
- Willingness and ability to quickly adapt to new environments and learn new technologies
- Experience in Data Analysis, Data Migrations, Data Cleaning, Transformation, Integration, Data Imports and Data Exports
- Hands on experience in optimizing the SQL Queries and database performance tuning in SQL Server database
- Performed data visualization with Tableau and generated dashboards to present the findings.
- Good knowledge on libraries and frameworks such as Pandas, NumPy, NLTK, Stanford NLP, Weka, Spacy in Text mining, social media analytics and big data engineering.
- Good working knowledge in Tableau visualization tool to creating dashboards and Visualize the data from various data sources.
- Having Good Knowledge on Hadoop Technologies like Hive, Sqoop, Oozie and its application Support.
- Importing and exporting data into HDFS and RDMS using Sqoop.
- Good knowledge of deep leaning neural networks such as CNN, RNN.

Core Knowledge & Skills

Languages:	Python, SQL, R(Basics)
Machine Learning Algorithms:	Linear & Logistic Regression Models, Decision Trees, Random Forest, Clustering Algorithms (K-means, Agglomerative), Time Series Forecasting's, Naive Bayes classifier,
Libraries:	Scikit-learns, TensorFlow, NumPy, Pandas, NLTK, Matplotlib, ggplot2, SPACY, Seaborn,
Statistical Methods & Techniques	Descriptive, Inferential methods, Hypothesis testing, Central limit theorem, ANOVA, T- test, Stats Model (OLS)
Database	Bigdata, MySQL, Oracle
Web Technology	HTML, CSS, BOOTSTRAP, Django
Operating Systems:	Linux, Windows.

Professional Experience

- Data Scientist - Dell EMC - August 2018 to till date.
- Data Analyst -Dell EMC - August 2017- July 2018
- Bigdata Support Engineer - Dell EMC - January 2017 to July 2017

Project summary

Project 4: Email, Ticket classification, sentiment Analysis and topic modeling.

Role: NLP Engineer

Technologies: Machine learning techniques, python, NLP, NLTK, SPACY

Duration: September 2020 to still going on

Responsibilities:

- Model was built with python, SPACY, scikitlearn, Pandas, Numpy, and NLTK.
- Communicate the results with operations team for taking best decisions and Collect data needs and requirements by Interacting with the other departments.
- Work with NLTK library to NLP data processing and finding the patterns.
- Categorize comments into positive and negative clusters from different social networking sites using Sentiment Analysis and Text Analytics.
- Ensure that the model has low False Positive Rate and Text classification and sentiment analysis for unstructured and semi-structured data.
- Used bag of words to build to build vocabulary using words in reviews
- Involved in the text preprocessing, data cleaning, Count Vectorizer and transform text into numerical vector data using a pipeline
- Vector space technique represents each word in language.

- Hands on experience in Natural Language processing, sentiment analysis, Symantec Analysis and Topic extraction.
- Used N-gram model and add and remove the stop words from the corpus to make Good prediction on text data.
- Used TF-IDF approach to give weights to words based on occurrences of a word
- Built a linear SVM classifier, naive Bayesian classifier to classify the text data.
- Analyzed the positive/negative sentiment for each given review to find out the like/dislike of the custom.
- Model was evaluated using metrics such as accuracy, precision and recall

Project 3: cluster Analysis

Role: Data Scientist

Technologies: Machine learning techniques, MySQL, Python, Pandas, Numpy,Matplotlib

Duration: July 2019 to August 2020

Responsibilities:

- The data employed involved 7 features that included Invoice No, StockCode , Description , Quantity, Invoice Date, Unit Price , Customer ID , Country and dataset having the Dim(4,000,7)with different set of customer's.
- Model was built with python, scikitlearn ,Pandas, Numpy, and matplotlib,Tableau.
- Worked with project team to understand the problem and business requirements.
- Converted all categorical data into continuous which will be used for modelling.
- Imported data into Python for exploring and understanding data
- Exploring the data and data structures for developing model
- Prepared data for creating training and test sets and data cleaning.
- Created a procedure to compute the cost of K-Means, this data was used to build a elbow chart to determine the optimum number of clusters to use.
- Model build on K-mean clustering, Agglomerative clustering and RFM model.
- Selected optimal number of clusters by elbow curve.
- Evaluate and selected the best k- cluster's by using Hopkins Statistics, Silhouette Analysis, Hierarchical Clustering, dendrogram.

Project 2: Churn prediction

Role: Data Scientist & Data Analyst

Technologies: Machine learning techniques, MySQL, Tableau, python, NumPy, pandas, matplotlib

Duration: August 2017 to June 2019

Responsibilities:

- Involving in the development of Python code, statistical/business model and its documentation.
- Identifying the relationships between multiple variables and using patterns in past data to shape insights.

- Data collection from various sources and validating the extracted data in view of continuous and categorical variables, along with checking for outliers and missing values.
 - Worked on outliers' identification with boxplot, K-means clustering using Pandas, Numpy.
 - Participated in features engineering such as feature intersection generating, feature normalize and Label encoding with Scikit-learn preprocessing.
 - Worked on outliers' identification with box-plot, K-means clustering using Pandas, Numpy.
 - Used Python 3.0 (numpy, scipy, pandas, scikit-learn, seaborn, NLTK) to develop variety of models and algorithms for analytic purposes.
 - Explored and analyzed the customer specific features by using dashboards in Tableau and by making visualizations using python
 - Used K folded technique to compare the accuracies across the models and pushed the results of the best models.
 - Visualized the patterns of customers through decision tree. Identified the prioritized variables by using Information Gain, Gini index, Entropy.
 - Analysed the classification report, accuracy score and ROC& AUC curve to gauge the performance of the models
 - Analyzed and grouped customers into different clusters based on customers purchase and historic data using techniques such as k-means clustering.
-
- Participated in features engineering such as feature intersection generating, feature normalize and Label encoding with Scikit-learn preprocessing.
 - Identified the TPR and FPR for the models.
 - Resampling Methods: (CROSS-Validation and BootStrapping) to increase the accuracy of an Machine Learning Algorithm.
 - Involving in Decision tree model building Tree pruning, setting constraints on tree Size to make good accuracy on test data.
 - Analyzed customer data for churn prediction using logistic regression, decision trees and Random Forests and done comparison on results
 - After an effective feature reduction, the recall of the model was increased by 6% and reach of the customer was increased by 11%

Project 1: AIG -RM (Risk management Analysis)

Role: Hadoop Support Engineer

Tools: Hadoop, Hive, Sqoop, Oozie, MySQL

Duration: January 2017 to July 2017

Responsibilities:

- Monitoring the CSM SIM job tracking, Analysis enhancement dashboard to better provided the services and solutions to the customer.
- Importing and exporting data into HDFS and RDMS using Sqoop.
- Involving in creating incident, service request ticket when the alert triggered over the dashboard.

- Managing long running jobs & huge cluster (Vcores/Memory) usage jobs and send the notification to respective user.
- Edge node and Hadoop space checking and notifying to the team to take necessary actions.
- Creating the CR Task as requested by resolver teams and following up on to get CR approval.
- Having Good Knowledge in writing Map Reduce jobs in Hive.
- Monitoring the scheduling jobs in HDFS by using Oozie.
- Managed and reviewed Hadoop log files.

Certifications

- **Associate - Data Science and Big Data Analytics v2** Completed from Dell Technologies proven professional.

Academic Qualification

- MCA (Master of computer applications) from JNTU University, Anantapur, 2015 passed out

Declaration

- I hereby declare that all the above furnished information is true to the best of my knowledge.

Place: Bangalore

Suresh kumar. G