**Sowan**

**Phone- 214 227 2114 RESUME**

**Email-** [**kush.m441@outlook.com**](mailto:kush.m441@outlook.com)

**PROFESSIONAL SUMMARY**

* About 5 years of experience in Data Science/Machine learning with large datasets of structured and unstructured data, Predictive modelling, Data analysis, Data acquisition, Data validation and Data visualization.
* Hands-on experience with  algorithms such as Regression Analysis, Clustering, Boosting, Classification, Principal Component Analysis and Data Visualization Tools.
* Data scientist with proven expertise in Data Analysis, Machine Learning, and Modeling.
* Experience in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, K-Means Clustering and Association Rules.
* Experience in applying predictive modeling and Machine Learning  algorithms for analytical reports.
* Experience using technology to work efficiently with datasets such as scripting, Data cleaning tools, statistical software packages.
* Developed predictive models using Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Cluster Analysis, and Neural Networks.
* Very Strong in Python, statistical analysis, tools, and modeling.
* Experienced in Machine Learning  and Statistical Analysis with Python Scikit-Learn.
* Strong programming skills in a variety of languages such as Python, R and SQL.
* Experience implementing Machine Learning back-end pipeline Spark ML-lib, Scikit-learn, Pandas, NumPy.
* Working knowledge of extract, transform, and Load (ETL) components and process flow using Talend
* Experience with AWS cloud services EC2, S3.
* Experience with Building and implementing architecture roadmaps for next generation Artificial Intelligence solutions for clients.
* Experience working with Agile and Waterfall models, dealing with sprints and resolving issues within each story of a sprint.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Big Data Technologies** | Hadoop, Hive, HDFS, MapReduce, Pig, Kafka. |
| **Machine Learning** | Regression, Polynomial Regression, Random Forest, Logistic Regression, Decision Trees, Classification, Clustering, Association, SVM, K-Nearest Neighbours (K-NN). |
| **BI Tools** | Tableau, Tableau server, Tableau Reader, SAP Business Objects, SAP Business Intelligence, Amazon Redshift |
| **Packages** | pandas, numPy, seaborn, sciPy, scikit-learn, Beautiful Soup, matplotlib, ggplot, NLTK |
| **Languages** | Python, R, C, C++, Java 8 |
| **Databases** | SQL, MySQL, PostgreSQL, Hive, MS Access, Mongo DB |
| **Reporting Tools** | MS Office (Word/Excel/Power Point/ Visio), Tableau, Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0. |
| **Version Control Tools** | GitHub. |
| **Operating System** | Windows, Linux, Unix |

**PROFESSIONAL EXPERIENCE**

**Client: Verizon Wireless, Dallas, TX Jul 2018 – Present**

**Role: Data Scientist**

***Project Description*:** Verizon Wireless is an American telecommunication giant which offers wireless products and services. The objective of the project was to determine customer satisfaction by performing sentiment analysis on twitter data and recommend marketing approaches based on customer consuming behavior to the management.

**Roles & Responsibilities:**

* Determine customer satisfaction and help enhance customer experience **using NLP.**
* Worked with NLTK library to NLP data processing and finding the patterns.
* Worked on **Text Analytics, Naive Bayes, Sentiment analysis, creating word clouds** and **retrieving data from** Twitter and other social networking platforms.
* Categorize comments into positive and negative clusters from different social networking sites using Sentiment Analysis and Text Analytics.
* Implemented technologies in NLP such as Noise Removal, Lemmatization, Stemming, Tokenizing, POS tagging, Bag of Words, Topic Modelling, TF-IDF, Word2Vec etc.
* Recommended and evaluated marketing approaches based on quality analytics of customer consuming behavior.
* Used AWS S3 service to store daily data from Twitter feeds in a text format using Python Boto library.
* Conceptualized the most-used product module (Research Center) after building a business case for approval, gathering requirements and designing the User Interface
* A team member of Analytical Group and assisted in designing and development of statistical models for the end clients. Coordinated with end users for designing and implementation of e-commerce analytics solutions as per project proposals.
* Conducted market research for client; developed and designed sampling methodologies, and analyzed the survey data for pricing and availability of clients' products. Investigated product feasibility by performing analyses that include market sizing, competitive analysis and positioning.
* Successfully optimized codes in **Python** to solve a variety of purposes in **data mining** and **machine learning in Python.**
* Facilitated stakeholder meetings and sprint reviews to drive project completion.
* Successfully managed projects using **Agile development methodology.**
* Project experience in **Data mining**, segmentation analysis, business forecasting and association rule mining using Large Data Sets with **Machine Learning.**

**Environment**: Python, MATLAB, MongoDB, exploratory analysis, Naïve Bayes K-Means Clustering, Hierarchical Clustering, Spark (MLlib, PySpark), Tableau, SAS, Hadoop 2.7, OLTP, Random Forest, OLAP, HDFS, ODS, NLTK, SVM, JSON, XML and MapReduce.

**Client:** Pacific Gas and Electric Company, San Francisco**,** CA **Aug 2017 - Jul 2018**

**Role: Data Scientist/Machine Learning Engineer**

***Project Description*:**

Development of a machine learning model to forecast the demand on the electric grid based on historical data as well as macroeconomic and demographic indicators. The objective was to help the executives make investment decisions on capacity expansion by predicting the short-term and long-term load on the system with reasonable accuracy.

**Roles & Responsibilities:**

* Performed intensive data pre-processing, feature engineering, feature scaling using python.
* Develop and implement innovative data quality improvement tools.
* Involved in Peer Reviews, Functional and Requirement Reviews.
* Developed project requirements and deliverable timelines; execute efficiently to meet the plan timelines.
* Creating and support a data management workflow from data collection, storage, analysis to training and validation.
* Involved with data analysis, primarily identifying data sets, source data, meta data, data definitions and data formats.
* Performed Data Cleaning, Feature Scaling, Feature Engineering using Pandas and Numpy libraries and applied Principle Component Analysis (PCA) for dimensionality reduction.
* Understanding requirements, significance of weld point data, energy efficiency using large datasets.
* Creating and support a data management workflow from data collection, storage, analysis to training and validation.
* Wrangled data, worked on large datasets (acquired data and cleaned the data), analyzed trends by making visualization tools (power BI, Tableau 9.0) using matplotlib and python.
* Understanding the business problems and analyzing the data by using appropriate Statistical models to generate insights.
* Knowledge of Information Extraction and NLP algorithms
* Developed NLP models for Topic extraction, Sentiment Analysis.
* Identify and assess available machine learning and statistical analysis libraries (including regressors, classifiers, statistical tests, and clustering algorithms).

**Environment:** Python, Machine learning, Pandas, SQL, Spark, AWS(S3/Redshift), Scikit-learn, Data Warehouse, Apache, Tableau.

**AMNIL Technologies, Nepal Nov 2014 – Jun 2017**

### Role: Data Analyst

**Responsibilities:**

* Involved in Analysis & Marketing Team to make business decisions.
* Involved with key departments to analyze areas and discuss the primary model requirements for the project.
* Documented methodology, data reports and machine learning model results and communicated with the Project Team Manager to share the knowledge.
* Well experienced in Normalization and De-Normalization techniques for optimum performance in relational and dimensional database environments
* Performed machine learning to estimate the probability of a new customer being classified as good or bad.
* Design develop and produce reports that connect quantitative data to give better insights.
* Involved in defining the source to business rules, target data mappings, data definitions.
* Built various **graphs** for business decision making using **Python matplotlib** library.
* Used **Python** library **Beautiful Soup** for **web scrapping** to extract data for building **graphs**.
* Performed troubleshooting, fixed and deployed many **Python bug** fixes of the two main applications that were a main source of data for both customers and internal customer service team
* Responsible for defining the key identifiers for each mapping/interface.
* Remain knowledgeable in all areas of business operations in order to identify systems needs and requirements.
* Document data quality and traceability documents for each source interface.
* Establish standards of procedures.
* Coordinated meetings with vendors to define requirements and system interaction agreement documentation between client and vendor system.

**Environment:** Python, R, Linux, Spark, Tableau, SQL Server 2012, Microsoft Excel, MATLAB, SQL, Scikit-learn, Pandas, XML, SQL Profiler, and Query Analyze.

**Education Details: Available upon Request**

**Reference: Available upon Request**