**Sushma Devulapalli**

**Data Scientist**

**Email:** [**dvlsushma201@gmail.com**](mailto:dvlsushma201@gmail.com)

**Ph #:**

**Professional Summary:**

* **7+ years** of hands on experience and comprehensive industry knowledge of **Machine Learning, Statistical Modeling, Data Analytics, Data Modeling, Data Analysis, Data Mining, Risk Management, & Natural Language Processing (NLP), Artificial Intelligence algorithms, Business Intelligence, Analytics Models (like Decision Trees, Linear & Logistic Regression, R, Python, SQL.**
* Experience in Machine Learning, Mathematical Modeling and Operations Research. Comfortable with **R, Python, MATLAB.**
* Expertise in **Data Analysis, Data Migration, Data Profiling, Data Cleansing, Transformation, Integration**, Data Import, and Data Export through the use of multiple ETL tools such as Informatica Power Center.
* Experience in working with **Regression** and Monte Carlo methods with Classification models for **Risk Management.**
* Proficient in **Machine Learning, Data/Text Mining, Statistical Analysis** & Predictive Modeling.
* Expertise in data acquisition, storage, analysis, integration, predictive modeling, logistic regression, decision trees, data mining methods, forecasting, factor analysis, cluster analysis, **ANOVA** and other advanced statistical techniques.
* Experienced in building data models using machine learning techniques for Classification, Regression, Clustering and Associative mining.
* Very good experience and knowledge in provisioning virtual clusters under **AWS** **cloud** which includes services like **EC2.**
* Proficient in data visualization tools such as **Tableau, Python Matplotlib, R** Shiny to create visually powerful and actionable interactive reports and dashboards.
* Proficiency and understanding of supervised and unsupervised machine learning algorithms like Random forest, SVM, Clustering and neural networks.
* Experience in Deep learning algorithms, TensorFlow, Keras.
* Hands-on-Experience in designing and developing NLP models for sentiment Analysis.
* Proficient in working with several python and R libraries including Pandas, NumPy, matplotlib, Scikit-learn, SciPy, GGplot, NLTK, TwitterR, ROAuth, Beautiful-soup, python-twitter, PyMySQL and sqlalchemy.
* Experience in presenting data using traditional analytic/visualization tools like Tableau, Excel.
* Knowledge of BI related principles such as ETL, data modeling and data warehousing
* Excellent Tableau Developer, expertise in building, publishing customized interactive reports and dashboards with customized parameters and user-filters using **Tableau (9.x/10.x)**
* Experienced in **Agile** methodology and SCRUM process.
* Strong business sense and abilities to communicate data insights to both technical and nontechnical clients.
* Working knowledge of A/B testing and web analytics tools.

**Technical Skills:**

|  |  |
| --- | --- |
| **Programming** | Python, C# Java, C++, C. |
| **Machine Learning** | Regression, Classification, Clustering, Recommendation Systems, NLP, Web Scraping, Data Analysis and Data Visualization. |
| **Algorithms** | KNN, Decision Trees, SVM, Logistic Regression, XGBoost, Gradient Descent. |
| **Statistics** | Bayes Theorem, Regression |
| **Frameworks and Libraries** | Django, Numpy, Scikit-learn, Pandas, Matplotlib, Seaborn, Tensorflow/Keras. |
| **Database** | SQL Server, Oracle |
| **Data visualization Tool** | Tableau |
| **Cloud Platform** | Watson Studio |

**Professional Experience:**

**Infosys Atlanta, GA**

**Role: Data Scientist Feb2021-TillNow**

**Responsibilities**

* Developed anomaly detection model to predict Network Transaction attempts
* Contributed to the Fraud rules by predicting the chance of fraud happening with 94% accuracy
* Developed Machine Learning solution to Flag high risk transactions
* Strong understanding of customer needs and business strategy/process
* Performed Data cleaning, Eda, Feature Extraction, Model Training and prediction.
* worked on generating unit test execution report through pytest html report generator
* Integrated the ServiceNow with SNS to create automatic ticket for specific failure.
* Experience working with AWS Sage Maker, S3.
* Experience in Data Wrangling, Statistical Modeling, Data Visualization.

Client: Telecom, Alpharetta, GA **Jun 2020 – Jan 2021**

**Role: Data Scientist**

**Responsibilities:**

* Manipulated large data sets and used them to identify trends and reach meaningful conclusions to inform strategic business decisions.
* Performed data collection, preprocessing and analysis.
* Presented information using data visualization techniques.
* Used NLP to classify text documents into more categories.
* Improved user experiences for products.
* Provided ideas for brand improvement, by analyzing insights avoided customer churn.
* Contribute to the creation of fraud rules by providing analytical data to support decisions.
* Identified fraud patterns through the monitoring of high-risk countries and merchants.
* Report risk information and findings to the immediate supervisor in a timely manner.
* Actively participated in team meetings; share information and make suggestions for improvements to fraud operations.
* Developed advanced analytical models and computational solutions using large-scale data manipulation and transformation, statistical analysis, machine learning, visualization.
* Developed custom data models and algorithms to apply to data sets. Used predictive modeling to increase and optimize customer experiences, revenue generation, ad targeting and other business outcomes.
* Setup storage and data analysis tools in Amazon Web Services cloud computing infrastructure using MySql Server.
* Involved in developing, scalable, fault tolerant systems on AWS.
* Used Data Quality Validation techniques to validate Critical Data Elements (CDE) and identified various anomalies.
* Participated in all phases of Data-Mining, Data-collection, Data-Cleaning, Developing-Models, Validation, Visualization and Performed Gap Analysis.
* Interaction with Business Analyst, SMEs and other Data Architects to understand Business needs and functionality for various project solutions
* Collaborate with external partners such as agencies to assist with data collection, reporting and A/B test development.
* Worked on creating ETL pipelines using python, different AWS services like Dynamodb, Glue jobs, step functions etc.
* Queried and aggregated data from Amazon Redshift to get the sample dataset.

**Client: MVB Bank, Fairmont, WV Jun 2019 – May 2020**

**Role: Data Scientist**

**Responsibilities:**

* Analyzed data using Python, PySpark, SparkSQL in order to do real time stream analytics.
* Documented logical, physical, relational and dimensional data models.
* Designed the Data Marts in dimensional data modeling using star and snowflake schemas.
* Predicting results using linear regression model with an error of 1%, using the statistical analytical tools and algorithms and helping in integrating the results into their sales and operations tools.
* Using NLP analyzed person's sentiment from his post
* Used NLP for spam detection and text summarization.
* Built Tableau dashboards that tracked the pre and post changes in customer behavior post campaign launch.
* Modeling and exponential smoothening for multivariate time series data.
* Developed segments using K-means, Gaussian mixture techniques.
* Developed a machine learning system that predicted purchase probability at a particular offer based on customer’s real time location data and past purchase behavior.
* Used Pandas, NumPy, seaborn, SciPy, Matplotlib, Scikit-learn, NLTK in Python for developing various machine learning algorithms and utilized machine learning algorithms such as linear regression, multivariate regression, naive Bayes, Random Forests, K-means, & KNN for data analysis.
* Implemented Normalization and Standardization preprocessing techniques for models like regression and KNN to reduce loss.
* Performed Grid search for a better choice of hyper parameters.
* Applied Image data processing techniques like image re scaling, Shearing, Contrast, Color balancing to get meaningful data from image.
* Performed Source System Analysis, database design, data modeling for the warehouse layer using MLDM concepts and package layer using Dimensional modeling.
* Implemented Predictive analytics and machine learning algorithms to forecast key metrics in the form of designed dashboards on to AWS (S3/EC2) and Django platform for the business.
* Created tables, sequences, synonyms, joins, functions and operators in Various Databases.
* Performed Data Analysis and Data Profiling and worked on data transformations and data quality rules.
* Involved in Creating A/B test plans and manage implementation

**Client: Mud Pie, Stone Mountain, GA Oct 2017 – May 2019**

**Role: Data Scientist**

**Responsibilities:**

* Involved in building database Model, APIs and Views utilizing Python,in order to build an interactive web based solution.
* Using GIT version control tool to coordinate team-development.
* Gathered reviews about the products and made suggestions for the improvement of the products using NLP.
* Developed the required XML Schema documents and implemented the framework for parsing XML documents.
* Wrote Python routines to log into the websites and fetch data for selected options. Used Python modules such as requests, urllib, urllib2 for web crawling.
* Involved in daily Scrum meetings and follows agile methodology.
* Created a Python based GUI application For Freight Tracking and processing.
* Developed automated testing framework for command-line based tests on Linux using Objected Oriented Perl and for selenium-based tests using Python.
* Created database using MySQL, wrote several queries to extract data from database.
* Executed Data Analysis and Data Visualization on survey data using Tableau Desktop as well as Compared respondent's demographics data with Univariate Analysis using Python (Pandas, NumPy, ScikitLearn, and Matplotlib).
* Developed Star and Snowflake schemas based dimensional model to develop the data warehouse.
* Designed Context Flow Diagrams, Structure Chart and ER- diagrams.
* Worked with SME's and other stakeholders to determine the requirements to identify Entities and Attributes to build
* Generated ad-hoc SQL queries using joins, database connections and transformation rules to fetch data from legacy SQL Server database systems.
* Created Logical and Physical data models with Star and Snowflake schema techniques using Erwin in Data warehouse as well as in Data Mart.
* Performed data analysis and data profiling using complex SQL on various sources systems including Oracle.
* Involved in Data Analysis, Data Validation, Data Cleansing, Data Verification and identifying data mismatch.
* Designed data model, analyzed data for online transactional processing (OLTP) and Online Analytical Processing (OLAP) systems.
* Experience in Agile Methodologies, Scrum stories and sprint experience in a python based environment along with data analytics, data wrangling.

**Client: Cross Country Healthcare, Boca Raton, FL Jan 2015 – Sep 2017**

**Role: Data Analyst**

**Responsibilities:**

* Work with users to identify the most appropriate source of record and profile the data required for sales and service.
* Document the complete process flow to describe program development, logic, testing, and implementation, application integration, coding.
* Involved in defining the business/transformation rules applied for sales and service data.
* Define the list codes and code conversions between the source systems and the data mart.
* Worked with internal architects and, assisting in the development of current and target state data architectures
* Worked with project team representatives to ensure that logical and physical ER/Studio data models were developed in line with corporate standards and guidelines
* Involved in defining the source to target data mappings, business rules, business and  data definitions
* Responsible for defining the key identifiers for each mapping/interface
* Responsible for defining the functional requirement documents for each source to target interface.
* Document, clarify, and communicate requests for change requests with the requestor and coordinate with the development and testing team.
* Involved in configuration management in the process of creating and maintaining an up-to-date record of all the components of the development efforts in coding and designing schemas
* Developed the financing reporting requirements by analyzing the existing business objects reports
* Interact with computer systems end-users and project business sponsors to determine, document, and obtain signoff on business requirements.
* Ran SQL queries to test and validate data in database (SQL Server, DB2).
* Developed SQL queries /scripts to validate the data such as checking duplicates, null values, truncated values and ensuring correct data aggregations within an ETL process testing cycle.
* Utilized Informatica toolset (Informatica Data Explorer, and Informatica Data Quality) to analyze legacy data for data profiling.
* Responsible in maintaining the Enterprise Metadata Library with any changes or updates
* Document data quality and traceability documents for each source interface
* Establish standards of procedures.
* Generate weekly and monthly asset inventory reports.
* Evaluated data profiling, cleansing, integration and extraction tools(e.g. Informatica)
* Coordinate with the business users in providing appropriate, effective and efficient way to design the new reporting needs based on the user with the existing functionality.

Education:

Masters in Computer Science from Osmania University , Hyderabad, India.

Bachelors in Mathematics and Statistics from Osmania University ,Hyderabad, India .