**Deepak Vidhur Damera**

Data Engineer

**PROFESSIONAL SUMMARY**

* 5+ years of experience in data-driven analytics and developmentthat has impacted business strategy, product marketing and operations.
* Identified business needs; Processed, filtered and analyzed data to **optimize** business operations and products.
* Brought into limelight and monitored **KPIs** that are significant to building and improving marketing strategies.
* Developed different Statistical & Machine Learning solutions to various business problems and generated data visualizations using Python, **ggplot** and **Matplotlib**.
* Experience with Data Science toolkits, such as **R Studio, Jupyter Lab**.
* Gathered insights from large data sets.
* Experience in R (Data Exploration, Manipulation, Summarization and Visualization & Analysis).
* Built various statistical models to solve the complex business problems.
* Knowledge of statistical techniques such as Descriptive Modelling, Linear & Non-linear Models, classification and Data reduction techniques, Predictive Analysis.
* Strong foundations including coursework and professional experience in machine learning algorithms, such as **Decision trees, Naive Bayes, SVM & Random Forrest, KNN, CNN** etc.
* Developed a k-prototype model for identification outliers in multi type data in Python.
* Developed an algorithm in python that automated price forecasting and sent an e-mail.
* Applied classification algorithms using R to predict whether an email is spam.
* Built fuzzy matching algorithm using k-nearest neighbors to identify non-exact matching duplicates.
* Worked on Variable reduction techniques to remove the insignificant variables and built sustainable ensemble models and voting classifiers.
* Domain specific knowledge in Sentiment Analysis (NLP), Corporate Credit Risk & Finance.
* Strong experience in Data Analysis, Data Profiling, Data Cleansing & Quality, Data Migration, Data Integration.
* **Experienced Data Modeler** with strong conceptual, Logical and Physical Data Modeling skills, Data Profiling skills, Maintaining Data Quality, creating data mapping documents, writing functional specifications, queries.
* Expert in backend testing to verify Data integrity and Data validation of client server and web-based applications using **SQL**.
* Experience in creating views for reporting, involves complex SQL queries with sub-queries, inline views and multi table joins, with clause and outer joins as per the functional needs.
* Good knowledge in Teradata, Business Objects, Crystal Reports, Ad hoc reports, MS Excel.
* Working experience in agile delivery environments and all phases of Software Development Life Cycle (**SDLC**).
* Strong analytical skills with the ability to collect, organize, analyze, and disseminate significant amounts of information to gather insights with attention to detail and accuracy.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Scripting Languages** | Python (NumPy, SciPy, Pandas, matplotlib)  R (ggally, dplyr, tidyeval), SAS  Frameworks: Tensorflow, PyTorch |
| **Machine Learning** | Regression, Decision trees, Support Vector Machines, XGboost, PCA,  K-fold validation, NN’s |
| **Databases** | HDFS, SAP Hana, MS Sql Server, PostgreSQL |
| **Cloud technologies** | AWS, Azure |
| **Operating System** | Windows, Unix |
| **Tools & Software** | Teradata SQL Assistant, TOAD, MS Office, BTEQ. |
| **Programming Languages** | R, SQL, PL/SQL, UNIX shell Scripting, C#, Python |
| **Execution Strategies** | Agile, Scrum, Lean Six Sigma, Ralph Kimball and Bill Inmon data warehousing methodology |

**WORK EXPERIENCE**

**Verizon, Tampa, FL Oct. 2018 – Nov. 2019**

**Data Engineer**

**Responsibilities:**

* Worked with senior managers and stakeholders to identify business requirements.
* Effectively communicated the achieved results and exceeded regional operational targets.
* Developed different Statistical & Machine Learning solutions to various business problems and generated data visualizations using Python ggplot and Matplotlib.
* Worked with Text analytics, extracting data using Google’s OCR (**tesseract**), parsing data using various **NLP** tools, revealing insights from text documents using different techniques like Latent Dirichlet allocation, K-Means, LSA and Sequence to Sequence modelling. Utilized analytics involving large data sets to support R&D.
* Enhanced experience in Pandas, Numpy, **Scikit-learn**, Matplotlib and NLTk in python for developing different Machine Learning models.
* Good knowledge in Tensorflow to implement Deep learning models, CNN for image recognition, Recurrent Neural Networks for Language translation.
* Gained concrete experience in Big Data technologies like **Hadoop, PySpark**.
* Achieved commendable results by creating and presenting**ensemble** models.
* Worked on Variable reduction techniques to remove the insignificant variables and built regression model.
* Worked on data cleaning with missing values and NA values in data sets.
* Advocated for the use of analytics where needed.
* Investigated and implemented new scientific analysis and methodologies to support big data analytics efforts
* Worked on the data for normality checking for performing analysis and outlier detection.
* Implemented different classification approaches for business decision making.
* Explained statistical concepts to non-technical associates in an abstract and comprehendible manner.

**ENVIRONMENT:** Python, SQL, Regression, Cluster analysis, Scala NLP, Spark, Kafka, MongoDB, Logistic Regression, Hadoop, Hive, Teradata, random forest, OLAP, Azure, HDFS, ODS, NLTK, SVM, JSON, Tableau, XML, tensorflow, MapReduce, AWS, tesseract.

**Leidos, McLean, VA Oct. 2017 – Sep 2018**

**Data Scientist/Engineer**

**Responsibilities:**

* Collected, studied, and interpreted large datasets; conducted reports; performed accurate, successful data management.
* Developed and implemented new forecasting models which increased company productivity and efficiency.
* Participated in monthly meetings with executives, provided information on the progress.
* Participated in continuous process improvements and worked across cross-functional technical teams to enhance insights board capabilities.
* Designed and increased the accuracy of several credit risk models that assess user profiles using text mining, sentimental analysis, NTP for fraud prevention.
* Hands-on experience developing supervised and unsupervised machine learning algorithms (regression, decision trees/random forest, neural networks, feature selection/reduction, clustering, parameter tuning, etc.),
* Experience in evaluating and making decisions around the use of new or existing methodologies, process, and tools.
* Advanced knowledge in model evaluation, tuning and performance, operationalization and scalability of scientific techniques and establishing decision strategies.
* Worked with cross-functional teams and multi-dimensional data. Automating the data flow using Control-M.
* Extensive involved in Data Analysis, Data Cleansing, Requirements gathering, Data Mapping, Functional and Technical design docs, and Process Flow diagrams.
* Involved in extensive DATA validation using SQL queries and back-end testing.
* Leveraging Driver tables to pull data in Teradata, worked on Django API's for accessing the database.
* Connected to RedShift through Tableau to extract live data for real time analysis.
* Created Tableau Dashboard for the top key Performance Indicators for the top management by connecting various data sources like Excel, Flat files and SQL Database.
* Created Tableau scorecards, dashboards using stack bars, bar graphs, scattered plots, geographical maps, Gantt charts using show me functionality.
* Developed Views and Templates with Python and using Django's view controller and template language, Website interface is created.
* Designed and created backend data access modules using PL/SQL stored procedures and Oracle.
* Good experience in writing SQL **Queries** and implementing stored procedures, functions, packages, tables, views, Cursors, triggers.
* Experience in using collections in Python for manipulating and looping different user defined objects.
* Developed and executed User Acceptance Testing portion of test plan
* Analyzed user behavior to offer recommendation for financial products.
* Collaborated with the Operations and Technology Department on the development of new automated data management/analysis software which increased the overall productivity and cut unnecessary costs.
* The Data Science team contributed towards the company’s effort to optimize and successfully reached the target.

**ENVIRONMENT:** Python, SQL, Teradata, Django API’s, Tableau, Jira, Oracle DB, Control-M

**Sugar CRM, Durham, NC Jan. 2016 – Sep. 2017**

**Data Analyst**

**Responsibilities:**

* Responsible for gathering requirements from Business Analysts and Operational Analysts and identifying the data sources required for the request.
* Performed Data analysis on many ad hoc request's and critical projects through which some of the critical business decisions will be made in marketing and resource allocation.
* Worked on Data Verifications and Validations to evaluate the data generated according to the requirements is appropriate and consistent.
* Optimized the data environment in order to efficiently access data Marts and implemented efficient data extraction routines for the delivery of data.
* Designed and developed weekly, monthly reports related to the marketing and financial departments using Teradata SQL.
* Responsible for business requirements gathering to build the enterprise Data warehouse
* Used Agile (SCRUM) methodology for project management through all the phases of SDLC (Software Development Life Cycle), created USER STORIES
* Understanding the various Business Processes involved in building the Data model and perform analysis
* Consolidating data from different sources like DB2, SQL Server & Oracle for data manipulation and data cleansing purpose using SAS Data Integration Studio 4.2 and SAS Enterprise Guide.
* Performing Gap Analysis and working on Data Mapping between the Warehouse, Data mart and ETL
* Write SQL queries to load Oracle data and text files into SAS data warehouse.
* Experience with troubleshooting and Tableau Analysis.
* Experience in tableau created Rapid Analytics and Reporting.
* Experience with Tableau for Data Acquisition and data visualizations for client and customers.
* Performed train handling simulation and analysis on clearing run data in preparation for project engineer review of performance issues for future field run validation. Developed a set list of criteria required for proper simulation execution to reduce rework and overall time devoted for completion of a simulation.
* Developed models for various profiling and customer segmentation for business and personal banking.
* Streamline the various reports for automation purpose using Base/SAS and Macros.
* Administration activities using SAS Management Console.
* Working with key stakeholders to document business processes and requirements.
* Used Star Schema and Snowflake schema methodologies in building and designing the physical data model.
* Developed and edited seven process documents, four validation documents, and one-track database format document for Trip Optimizer track database tool usage and development.
* Analyze, design, code, test, implement and a support data warehousing extract programs and end-user reports and queries.
* Developed different Statistical & Machine Learning solutions to various business problems and generated data visualizations using Python ggplot and Matplotlib.

**ENVIRONMENT:** Python, SQL, SAS, random forest, SQL server, SVM, JSON, Tableau