Dharmil Nitin
Data Engineer/Data Scientist

+1 (732) 746-5316   peter@ameripros.net

https://www.linkedin.com/in/dharmil1234

Professional Summary:

- Over 8+ years of professional experience as a Data Scientist/Data Engineer with a broad range of expertise in web-based applications. Data Scientist working across Finance and Telecom industries.
- Activates strong analytical skills to investigate trends in large amounts of data and formulate conclusions based on findings.
- Have Extensive Experience in IT data analytics projects, Hands on experience in migrating on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query, Cloud Data Proc, Google Cloud Storage, Composer.
- Excel at cultivating, managing, and leveraging client relationships to foster extended engagement and business opportunities.
- 5+ years of industry experience working with large database management systems and writing SQL queries and procedures.
- Working knowledge of Data Visualization including producing tables, graphs, listings using various procedures and tools such as Tableau and Power BI.
- 3+ years of experience working with Python and R to manipulate data for data loading and extraction and worked with blend of Python and R libraries for Data Management and Analysis.
- Strong background & hands - on in Data Science, Big Data, data structures, statistics, algorithms like Regression, Classification etc.
- Proficient in using AWS cloud services, working experience with ETL PySpark and SSIS.
- Proficient in relational Data modeling, Dimensional data modeling, logical/Physical Design, ER Diagrams and OLTP and OLAP System Study and Analysis.
- Proficient at data visualization tools such as Tableau, Power BI, Python Matplotlib and Seaborn.
- Experience as Big Data/Hadoop developer in Hadoop Ecosystem using Spark for complex business problems, which involves real time effective analysis and processing of semi-structured and unstructured data.
- Work experience in various environments such as: Python, PyCharm, Spark /spark, Eclipse, SQL, GitHub, AWS, Linux.
- Worked on Text Mining and Sentimental analysis for extracting unstructured data from various social Media platforms like Facebook, Twitter, and Reddit.
- Worked with the Relational database systems like MySQL, MSSQL, PL/SQL, Oracle, DB2, SQL Server, RDBMS, NoSQL database systems like Redis, PostgreSQL, MongoDB
- Experience in Full SDLC cycle which involves architecture, analysis, design, development, testing, Implementation, deployment, Enhancements, and production support using Agile and Waterfall Methodologies.
- Strong team player with good communication, analytical, presentation and interpersonal skills.
- Experience working with data lake is a system or repository of data stored in its natural/raw format.
- Working with a data lake is usually a single store of all enterprise data including raw copies of source system data and transformed data used for tasks such as reporting, visualization, advanced analytics, and machine learning.
- Experienced in Non-relational database such as MongoDB 3.x,
- Expert at distilling questions, wrangling data, and driving decisions with data analytics.
- Knowledge on time series analysis using AR, MA, ARIMA, GARCH and ARCH model.
- Can work parallelly in both GCP and Azure clouds coherently.
- Experienced in Agile methodology and SCRUM process.
- Excellent Tableau Developer, expertise in building, publishing customized interactive reports and dashboards with customized parameters and user - filters using Tableau (9.x/10.x).

Technical Skills:

Languages	Python, R, SQL, PL/SQL, SAS, T-SQL
Python Libraries	Beautiful Soup, NumPy, SciPy, matplotlib, python - twitter, Pandas, data frame, urllib2, SVN, Seaborn, NLP, Matrix Factorization.
Cloud	AWS, Microsoft Azure, GCP, IBM Cloud, Bluemix
Databases	My SQL, NoSQL, MS- SQL, SQL Server, MS Access, Oracle MongoDB and DB2, Hive, Impala, Pig, Spark SQL, HDFS, HBase, Teradata, Netezza, MongoDB,

	Cassandra, SAP Hana
Methodologies	SDLC, Agile, Waterfall, Scrum.
Operating systems	Windows, Linux/Unix (via PuTTY client), PowerShell, HTML, XML, CSS
Reporting and Visualization Tools	Tableau, Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0, Power BI, Google Analytics, MS Office (Word/Excel/Power Point/ Visio), Matplotlib, Seaborn, ggplot2
Libraries	Pandas, NumPy, Matplotlib, Sklearn, folium, Ggplot2, Caret.
Big Data Technologies	Hadoop, Hive, HDFS, MapReduce, Pig, Kafka, PySpark
Bug Tracking Tools	JIRA, SPSS, Alteryx, SAS

Professional Experience:

Data Scientist/Data Engineer
BHG Financial – Atlanta, GA   **Nov 21 – Till Date**

Responsibilities:
- Worked as a Data Modeler/Analyst to generate Data Models using Erwin and developed relational database system. A highly immersive Data Science program involving Data Manipulation & Visualization, Web Scraping, Machine Learning, SQL, GIT, Unix Commands, NoSQL, MongoDB.
- Provide expertise and recommendations for physical database design, architecture, testing, performance tuning and implementation.
- Transformed Logical Data Model to Erwin, Physical Data Model ensuring the Primary Key and Foreign Key relationships in PDM, Consistency of definitions of Data Attributes and Primary Index Considerations.
- Collaborated with Database engineers to implement ETL process.
- Experience in GCP Dataproc, GCS, Cloud Functions, BigQuery.
- Experience in moving data between GCP and Azure using Azure Data Factory.
- Extensively worked on Data Modeling tools Erwin Data Modeler to design the data models.
- Setup storage and data analysis tools in Amazon Web Services cloud computing infrastructure.
- Data was visualized using different visualization (scatter plot, box plots, and histograms) techniques from ggplot2 package in R.
- Designed the physical model for implementing the model into theoracle9i physical database.
- Involved with Data Analysis Primarily Identifying Datasets, Source Data, Source Meta Data, Data Definitions and Data Formats
- Used Cloud Shell SDK in GCP to configure the services Data Proc, Storage, Big Query.
- Application of various machine learning algorithms and statistical modelling like Logistic Regression, Decision tree, SVM, to identify Volume using scikit-learn package in R & Python.
- Improve efficiency and accuracy by evaluating model in R.
- Designed tables and implemented the naming conventions for Logical and Physical Data Models in Erwin 7.0. Used R to manipulate data, develop and validate quantitative models.
- Cleansed the data by eliminating duplicate and inaccurate data in R and Python.
- Implementing Neural network for Large-scale adoption of Electronic Health Records (EHR) using H2O, Tensor Flow, OpenCV, and Keras.
- Solution architecting BIG Data solution for Projects & Proposal using Hadoop, Spark, ELK Stack, Kafka, Tensor flow.
- Built text analytics solutions with IBM Watson and Google Cloud which in turn saved analysts operational and processing time.
- Designed logical and physical data models for multiple OLTP and Analytic applications.
- Extensively used the Erwin design tool &Erwin model manager to create and maintain the Data Mart.
- Performance tuning of the database, which includes indexes, optimizing SQL statements, monitoring the server.
- Utilized Power BI (Power View) to create various analytical dashboards that depicts critical KPIs such as legal case matter, billing hours and case proceedings along with slicers and dicers enabling end-user to make filters.
- Utilized Power Query in Power BI to Pivot and Un-pivot the data model for data cleansing and data massaging.
- Wrote simple and advanced SQL queries and scripts to create standard and Ad hoc reports for senior managers.
- Collaborated the data mapping document from a source to target and the data quality assessments for the source data.

- Created S3 buckets and managed roles and policies for S3 buckets. Utilized S3 buckets and Glacier for file storage and backup on AWS cloud. Used Dynamo DB to store the data for metrics and backend reports.
- Worked with Elastic Beanstalk for quick deployment of services such as EC2 instances, Load balancer, and databases on the RDS on the AWS cloud environment.
- Used Java code to connect AWS S3 buckets by using AWS SDK, to access media files related to the application.
- Used Amazon Simple Workflow service (SWF) for data migration in data centers which automates the process and tracks every step and logs are maintained in S3 bucket.
- Designed and developed user interfaces and customization of Reports using Tableau and OBIEE and designed cubes for data visualization, mobile/web presentation with parameterization and cascading.
- Performed Data Analysis and Data Profiling and worked on data transformations and data quality rules.
- Created SSIS Packages using Pivot Transformation, Execute SQL Task, Data Flow Task, etc. to import data into the data warehouse.
- Developed and implemented SSIS, SSRS and SSAS application solutions for various business units across the organization.

Environment: SQL, GIT, Unix Commands, IBM Watson, Tensor flow, NoSQL, R/R studio, Python, ETl, GCP, Tableau, Mongo DB, SSIS, SSRS, SSAS, AWS, S3, EC2, RDS, SWF, Dynamo DB, Glacier, Erwin, Tableau, OBIEE.

Data Engineer/Data Scientist
Fitzco – Atlanta, GA   **Jul 20 – Oct 21**

Responsibilities:
- Building deep learning models to maximize cash collections and optimize workforce efficiency.
- Data Acquisition of the transactional claims level data and perform data mining to get meaningful data trends to perform Machine Learning modelling.
- Analyzing large volumes of structured and unstructured data using Python, R, Tableau and ML library such as Scikit Learn with an emphasis on evaluating data for anomaly detection, trend analysis, and data mining to support various business objectives.
- Writing complex SQL scripts for data acquisition. Perform flattening of long data sets and perform feature engineering on Claims data.
- Perform NLP analysis to determine which N number of topics are important for the disposition.
- Categorize each note into one of the N number of topics above using a document-matching NLP methodology.
- Developing and implementing various complex models using machine learning algorithms such as linear regression, classification, multivariate regression, Naive Bayes, Random Forests, K-means clustering, KNN, PCA, regularization and imputation.
- Performing a proper exploratory data analysis, used K-Means clustering technique to identify outliers. Dealt with unbalanced data by bootstrap resampling.
- Implemented cost effective and cutting-edge Data Science/Big Data Lab where I can easily prototype using the most advanced tools such as Jupyter Notebook, Docker Image, Python, R, NoSQL, RStudio, GPUs/CPUs.
- Developing optimized and scalable machine learning algorithms capable of performing predictive modeling using PyTorch, Spark, SparkML, Scikit-learn and other packages.
- Perform feature engineering to come up with effective input features for predictive models using one hot encoding.
- Utilizing principal component analysis and factor analysis to complete dimensional reduction of the data. Worked on ensemble Methods such as Bagging (Random Forest, Boosting (Ada-boost) & Stacking. Feature Engineering: Pearson Correlation, F-score, and Dimensionality Reduction: PCA, LDA.
- Identified, analyzed, and interpreted trends or patterns in complex data sets and provided business insights using LDA Topic Modeling.
- Responsible for analyzing patterns across multiple years. Feature engineered various metrics to track performance and make necessary improvements in statistical models. Tool/techniques leveraged include Python, R, TensorFlow, LDA, LSA and API's.
- Developed a prototype pipeline in Jupyter Notebook to pre-process data and to make user recommendations based on an ensemble of machine learning models.
- Designed and developed various interactive Intelligent dashboards in Tableau for decision making and various insights supporting business objectives.

- Interpreting business problems from a variety of business units, projects and designing technical solutions that may require work from various sub-disciplines of data science.
- Providing technical leadership on machine learning and data processing/analytics.
- Involved in Jira user stories creation, Kanban Process, Sprint Planning, Daily Standups, Sprint Demos, Story Grooming, Release Planning, and PO acceptance.

Environment: Anaconda, CRISP Data Mining, GPU/CPU Servers, Docker Containerization, Confluence, Bit Bucket, Jira, Kanban, Jupyter Labs, Tableau, Google Cloud Platform, Big Query, Netezza Data labs, Spark, HDFS, Hive, Pig, Linux, Python (Scikit- Learn/SciPy/NumPy/Pandas/Pytorch), R, MySQL, Microsoft SQL Server.

Data Scientist/Data Engineer
Symbiosis Technologies – Hyderabad, India   **Jun 16 – Jul 19**

Responsibilities:
- Cleaned and manipulated complex datasets to create the data foundation for further analysis and the development of key insights (MS SQL server, R, Tableau, Excel)
- Incorporated various machine learning algorithms and advanced statistical analysis like decision trees, regression models, SVM, clustering using scikit-learn package in Python.
- Worked on data pre-processing and cleaning the data to perform feature engineering and performed data imputation techniques for the missing values in the dataset using Python.
- Performed Exploratory Data Analysis (EDA) to visualize through various plots and graphs using matplotlib and seaborn library of python, and to understand and discover the patterns on the Data, understanding correlation in the features using heatmap, performed hypothesis testing to check significance of the features.
- Developed analytical approaches to answer high-level questions and provided insightful recommendations.
- Conducted various statistical analysis like linear regression, ANOVA, and classification models to the data for analysis.
- Involved in extracting customer's Big Data from various data sources (Excel, Flat Files, Oracle, SQL Server, MongoDB, Teradata, and log data from servers) into Hadoop HDFS
- Used pandas, NumPy, seaborn, SciPy, matplotlib, scikit-learn in Python for developing various machine learning algorithms.
- Assured data quality and data integrity, and optimized data collection procedures on a weekly and monthly basis
- Created Data Quality Scripts using SQL and Hive to validate successful data load and assured the quality of data.
- Worked on different data formats such as JSON, XML, CSV, .dat and exported the data into data visualization/ ETL platform.
- Evaluated model performance using techniques like R square, adjusted R square, confusion matrix, AUC, ROC curve, Root mean squared error etc.
- Incorporated, Developed, and applied metrics and prototypes that could be used to drive business decisions.
- Participated in ongoing research, and evaluation of new technologies and analytical solutions to optimize the model performance.
- Used problem-solving skills to find and correct data problems, applied statistical methods to adjust and project results when necessary.
- Worked across cross-functional teams to understand the data requirements and provided detailed analytical reports to accomplish the business decisions.

Environment: Model: Logistic Regression, Decision Trees, Random Forest, Python, MySQL, Machine Learning Libraries - NumPy, Pandas, Scikit, Stats-model

Data Scientist
Unisoft Global Services Pvt Ltd – Bengaluru, India   **Jun 14 – May 16**

Responsibilities:
- Utilized Python libraries like matplotlib, seaborn, NumPy, Plot.ly.
- Developed a layer of applications modules over the Python - Pandas library, delivered various data frame visualization tools, Data wrangling and cleaning using Python Pandas.
- Used python modules like Pandas, NumPy, Matplotlib, SciPy to perform data analysis.
- Created MS Access databases for day-to-day data analysis and statistics.

- Performed data cleaning and data processing using Python Data frames.
- Developed features and enhancements Python data structure libraries such as NumPy and Pandas. Extensively used matplotlib for generating graphical representation of data.
- Utilized Power BI to design multiple scorecards and dashboards to display information required by different departments and upper-level management.
- Managed Power BI Portal - Data refreshes / User management and set up gateway data refresh.
- Involved in data migration project to move data to AWS.
- Developed business process models of Waterfall to document existing and future business processes.
- Created MS Access databases for day-to-day data analysis and maintained Outline Agreements (Contracts) in SAP and Excel.
- Involved in developing and providing technical support of database objects like stored procedures, functions, triggers, views, and tables.
- Developed tools to automate some base tasks using SQL and Python.
- Developed monitoring application which captures the error related data and stores it in database.
- Collaborated within a team using an agile development workflow and widely accepted collaboration practices using Git.
- Involved in debugging the applications monitored on JIRA using agile methodology.

Environment: Python, matplotlib, seaborn, NumPy, Plot.ly, Pandas, Data Visualization, MS Access, Power BI, AWS, SAP, SQL, Agile, JIRA, Git, Data Analysis, Data Migration, SciPy

Educational details: Bachelor is in computer science – JNTU, Hyderabad, Telangana, India.