**DINESH NARSIMHA**

Rishitha.23g@gmail.com

+1 2142998373

[www.linkedin.com/in/dineshkgn |](http://www.linkedin.com/in/dineshkgn%20|) Tampa, FL

**Summary:**

* Having **8+**years of IT Industry experience and as Data Scientist with specialization in implementing advanced **Machine Learning** and **Natural Language Processing** algorithms upon data from diverse domains and building highly efficient models to derive actionable insights for business environments leveraging exploratory data analysis, feature engineering, statistical modeling and predictive analytics.
* Worked in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modelling and data visualization with large data sets of structured and unstructured data.
* Extensive experience in **Text Analytics**, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python.
* Had executive experience and performing IT roles for various industry leaders. Acquired a deep range of skills in:
* Proficient in data mining tools like **R, SAS, Python, SQL, Excel**, ecosystems Staff leadership and development
* Experienced with Machine Learning Algorithm such as **Logistic Regression, KNN, SVM, Random Forest, Neural Network, Linear Regression, Lasso Regression** and **K-Means.**
* Experienced working with data modelling tools like Erwin, Power Designer and ER Studio. Expertise in synthesizing Machine Learning, Predictive Analytics and Big data technologies into integrated solutions.
* Creating from scratch **Machine Learning** and **NLP** solutions for Big Data on top of **Spark** using **Python**.
* Skilled in Advanced Regression Modelling, Time Series Analysis, Statistical Testing, Correlation, Multivariate Analysis, Forecasting, Model Building, Business Intelligence tools and application of Statistical Concepts
* Proficient in data entry, data auditing, creating data reports & monitoring data for accuracy Ability to extract Web search and data collection, Web data mining, Extract database from website, Extract Data entry and Data processing.
* Developed **Spark, Python** for **regular expression(regex**) project in the **Hadoop/Hive** environment with **Linux/Windows** for big data resources. Used **K-Means** clusteringtechnique to identify outliers and classify unlabeled data.
* Experience with Data Analytics, Data Reporting, Adhoc Reporting, Graphs, Scales, PivotTables and OLAP reporting.
* Expert in data flow between primary DB and various reporting tools, Expert in finding Trends and Patterns within Datasets and providing recommendations accordingly.
* Utilized **Git** for version control and collaboration in my data science project at Bloomin Brands. Created and managed branches for feature development and bug fixing, and merged changes into the main branch using pull requests. Worked with team members to resolve merge conflicts.
* I’ve edited and maintained **JSON** files to configure data for use in e-commerce applications. Used **Visual** **Studio** **Code** and **JSON** validators to ensure proper syntax and structure of files.
* I’ve developed and implemented data analysis workflows on a **Hadoop** cluster using **HiveQL** and **Linux** command line tools. Conducted data cleaning, feature engineering, and predictive modeling on large datasets to perform analysis on customer behavior and improve business decision-making.
* Exploring DAG’s, their dependencies and logs using **Airflow** pipelines for Automation.
* Ability to use dimensionality reduction techniques and regularization techniques.
* Independently handle **Hadoop** administration in local and on cloud in Linux environment.
* Extracting and modelling datasets from verity of data sources like **Teradata** and **Snowflakes** for adhoc analysis and have fair understanding of **AGILE** methodology and practice.
* Working knowledge on Application design, architecture and development.
* Experienced in complete **SDLC** and **STLC** with end-user interaction for functional specification, system analysis, and unit regression testing; participated in system integration testing
* Experienced in working in a team environment to deliver on demand service; ability to deliver appropriate quality solutions under pressure; pro-active and strong analytical problem-solving skills.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Analytical Tools:** | Azure Databricks, SAS, Tableau, Google Analytics, Alteryx, PowerBI, Azure ML Studio, Excel |
| **Programming Languages:** | Python, PySpark, SQL, Java, C++, Java Script, C, NLP, R |
| **Big Data Ecosystems:** | Apache Hive, Apache Spark, HDFS, Map Reduce |
| **Databases & DBMS:** | SQL Server (SSMS), Oracle, Microsoft Azure Blob Storage, MySQL, Teradata SQL, Cassandra |
| **Applications:** | JupyterLab, BigQuery, PostgreSQL, MySQL, SQLite, GitHub, Command-Line, Tableau, Airflow |
| **Methods and Tools:** | Machine Learning (Supervised, Unsupervised, Neural Networks), Time Series Forecasting, Statistical Modeling, Scikit learn, TensorFlow, Pandas, NumPy, Seaborn, SQL Alchemy, JSON |

**EDUCATION:**

* **University of Cincinnati** | **Master’s Degree** - Computer and Information Science
* **Osmania University** | **Bachelor’s Degree** - Computer Science and Engineering

**COURSES:**

* Machine Learning, Artificial Intelligence, Intelligence Data Analysis, Database Management System, Operating Systems, Network Security, Advanced Algorithms, Image Processing.

**PROFESSIONAL EXPERIENCE:**

**Data Scientist| Bloomin’ Brands, Inc., Tampa, FL Aug 2018 – Present**

**Description:**

I developed various ML models in Spark/PySpark on Azure Databricks, including models to predict Customer Lifetime Value, churn likelihood, traffic/sales drivers for 900 restaurants, and customer segmentation based on purchase behavior. I partnered on Marketing Mix Models to estimate marketing effectiveness and created automated data pipelines in collaboration with business partners. I conducted customer data analysis in Python (scikit-learn), experimented with ensemble methods, and built neural networks using Python libraries like TensorFlow and Keras. Additionally, I performed NLP tasks for topic extraction and sentiment analysis, tracked visitor interaction using Google Analytics, and executed predictive analytics on AWS and Django platforms. I extensively used SQL Data Warehouse, SSIS, SSRS, and Hadoop for data wrangling and ETL processes.

**Responsibilities:**

* Developed ML model in Spark/PySpark (Azure Databricks) that prognosticates the Customer’s Lifetime Value and the customers who are most likely to churn based on their usage patterns, allowing for targeted retention campaigns.
* Developed a ML model that identifies the weekly drivers for traffic/sales of 900 restaurants in USA individually.
* Built ML model for **Customer Segmentation** based on a member’s purchase behavior and frequency patterns.
* Partnered with model team in building the **Marketing Mix Models** to quantitatively estimate the effectiveness of marketing elements.
* Developed and owned reporting on Loyalty customers acquisition, engagement, and retention metrics to track performance and marketing efficacy using **Alteryx** and **SAS** **EG**.
* Performed **NLP** based tokenization, lemmatization, vectorization and developed NLP models for **Topic Extraction, Sentiment Analysis**.
* Performed topic modeling, sentiment analysis and used word embeddings for clustering.
* Collaborated with business partners for data acquisition & validation and built automated data pipelines.
* Performed Customer Data Analysis in **Python (scikit-learn)** to target customers for marketing campaigns.
* Experimented with Ensemble methods to increase the accuracy of the training model with different Bagging and Boosting methods.
* Superintended usage of **Python NumPy, SciPy, Pandas, Mat plot, Stats** packages to perform dataset manipulation, data mapping, data cleansing and feature engineering. Built and analyzed datasets using **R and Python.**
* **Multi-layers Neural Networks** built in **Python Scikit-learn, Theano, Tensor Flow** and **keras** packages to implement machine learning models and export them into protobuf and performed integration job with client's application.
* Predominant practice of **Python** **Matplotlib** package and **Tableau** to visualize and graphically analyses the data. Data pre-processing, splitting the identified data set into Training set and Test set using other libraries in python.
* Loyalty Customer’s Life expectancy and the correlation in **Tableau**. (The data was cleaned and processed in excel and the visualization and correlation were done in Tableau)
* Tracked visitor interaction and traffic on the firm websites using **Google Analytics** and automated reports for tracking online traffic KPI’s in **Microsoft** **Power BI.**
* Worked on fraud detection analysis on payments transactions using the history of transactions with supervised learning methods.
* Built models using Statistical techniques like **Bayesian** **HMM** and ML classification models like **XG** **Boost**, **SVM** and, **Random Forest.**
* Predictive analytics and machine learning algorithms to forecast key metrics in the form of designed dashboards on to **Amazon Web Services** **(S3/EC2 CLOUD PLATFORMS)** and **Django** platform for the company's core business.
* Extensively dealt with Microsoft’s enterprise data wrangling and data forecasting in SQL Data Warehouse and conducted business performance reporting using **SSIS** and **SSRS** tools.
* Piping and processing massive data-streams in distributed computing environments such as Hadoop to facilitate analysis (ETL).
* Using **Airflow** to keep track of job statuses in repositories **MySQL** and Postgre databases.
* Assisted in conducting **A/B testing** across multiple channels that drive improvements in core menu pricing, execution, and experience.

**Environment:** Spark, PySpark, Azure Databricks, Python (scikit-learn, TensorFlow, Keras, NumPy, SciPy, Pandas, Matplotlib), Alteryx, SAS EG, Tableau, Microsoft Power BI, Google Analytics, AWS (S3, EC2), Django, SQL Data Warehouse, SSIS, SSRS, Hadoop, NLP techniques (tokenization, lemmatization, vectorization), Statistical techniques (Bayesian HMM, XGBoost, SVM, Random Forest), Airflow, MySQL, PostgreSQL.

**Data Scientist | Opera Solutions, Jersey City (Client: American Express, NY) Nov 2017 - Jul 2018**

**Description:**

I created financial reports and performed ad-hoc analysis using Apache Hive for large datasets. I developed Hive and Shell scripts for data processing, conducted data manipulation and statistical analysis in Python, and built and validated ML models, including regression and classification models using Keras. I worked on big data initiatives, collaborating with business partners to address data discrepancies in ETL pipelines and maximize ROI. I utilized AWS services (EC2, EMR, RDS, S3) for data storage and deployment, conducted EDA and data visualizations using Python and Tableau, and applied advanced statistical methods to identify trends and relationships. Additionally, I built an end-to-end data engineering pipeline using PySpark, created and managed A/B test plans, and optimized SQL queries for data extraction.

**Responsibilities:**

* Created financial reports and performed ad-hoc analysis from big data databases having 700 million records using Apache Hive.
* Designed and Developed **Hive** and **Shell Scripts**, Data Import/Export, Data Conversions and Data Cleansing.
* Performed data manipulation and cleansing with **Python** followed by statistical analysis such as **Hypothesis** **testing**.
* Applied models and regression, comparing various initial models, creating pipelines for data processing, and presenting reports to other teams within the company.
* Engaged with business groups and partners as part of an analytics team to analyze data discrepancies in ETL pipelines and come up with solutions to mitigate it. Built solutions to maximize program **ROI**.
* Responsible **for big data** initiatives and engagement including analysis, brainstorming, POC, and architecture.
* Worked with Cloud Services **AWS EC2, EMR, RDS, S3** to solve the data storage issue and work on deployment solution.
* Performed Exploratory Data Analysis and Data Visualizations using **Python**, and **Tableau**.
* Conducted statistical analysis of datasets using SQL queries and made creative implementations for operational streamlining.
* Aligned with Applied Artificial Intelligence in the Strategy and Analytics offering portfolio, built a Speech-to-text converter in a secured server with no internet access or API calls.
* Extracted features from audio files and built classification models using **Keras** to classify an audio file among different speakers and performed hyperparameter tuning.
* Tuned the performance of models by comparing different performance metrics like Confusion Matrix, precision, recall, ROC curve and accuracy using cross-validation.
* Built end to end Data Engineering pipeline using **PySpark** for an automotive manufacturer client.
* Created A/B test plans and managed implementation.
* Collaborated with data engineers, wrote and optimized **SQL** queries to perform data extraction from SQL tables.
* Determined trends and relationships in data by applying advanced statistical methods like T-test, hypothesis testing, ANOVA, Chi-Square test and Correlation analysis.
* Coordinated with the data scientist team and BA team to analyze on building a predictive model based on the requirements using various machine learning algorithms.

**Environment:** Apache Hive, Shell Scripts, Python (Pandas, NumPy, SciPy, Seaborn, Matplotlib), Keras, PySpark, AWS (EC2, EMR, RDS, S3), Tableau, SQL (queries, data extraction), Statistical methods (T-test, hypothesis testing, ANOVA, Chi-Square test, Correlation analysis)

**Data Analyst | Astral Technologies Inc. Charlotte, NC Jun 2016 – Nov 2017**

**Description:**

I performed extensive data cleaning and mining using Python to build predictive models that improved CTR by 30% and revenue by 20%. I implemented advanced text analytics using LDA, TF-IDF, and Word2Vec for buyer behavior analysis. I participated in all phases of data acquisition, cleaning, model development, validation, and visualization, generating statistical reports like Box Plots, Scatter Charts, and Heat Maps using Python libraries. I employed model validation metrics such as Confusion Matrix, Precision, Recall, and F1-Score, and used Grid Search for hyperparameter optimization. Additionally, I conducted visualizations and created reports using Tableau and Python Matplotlib, used Git for version control, and performed pre-research on big data techniques like Spark and Cassandra.

**Responsibilities:**

* Performed extensive data cleaning and data mining using python to build predictive models which improved CTR by 30% and revenue by 20%.
* Implemented advanced text analytics features like feature extraction using LDA method with TF-IDF vectorization and Word2Vec to analyze the new buyer behavior.
* Participated in all phases of data acquisition, data cleaning, developing models, validation, and visualization to deliver data science solutions.
* Implemented initial Exploratory Data Analysis (EDA) and generate statistical reports like Box Plot, Scatter Chart and Heat Map using Python NumPy, Seaborn, Pandas and Matplotlib.
* Performed Data Cleaning, Outlier Detection, Feature Scaling, and Feature Engineering using Python 3.6 and Python libraries such as Pandas, NumPy, SciPy and Seaborn.
* Employed Model validation Classification metrics which includes Confusion Matrix, Precision, Recall, Accuracy, F1-Score, Support and AUC for every model.
* Applied Grid Search for selecting best Hyperparameter to optimize best model and 10-fold cross validation to select models.
* Conducted visualization and creating reports by using Tableau and Python Matplotlib.
* Used Git for version control and source code management.
* Performed Pre-research on big data techniques such as Spark, Cassandra, NoSQL databases and assess the advantages and disadvantages of them.
* Assist in requirements gathering, design, project management, and UAT for all business intelligence projects.

**Environment:** Python (Pandas, NumPy, SciPy, Seaborn, Matplotlib), Text analytics (LDA, TF-IDF, Word2Vec), Model validation (Confusion Matrix, Precision, Recall, Accuracy, F1-Score, AUC), Grid Search, 10-fold cross validation, Tableau • Git • Big data techniques (Spark, Cassandra, NoSQL databases)

**PROJECTS**

**Automatic Blood Cell Counting *(University of Cincinnati)***

• Collaborated with research team and developed an application which automatically counts the number of blood cells present from an image sample using Image processing methods by **OpenCV**.

**Network Security and Android Malware (*University of Cincinnati)***

• Implemented Diffe-hellman for encrypted communication with the server, **RSA** and **zero knowledge proofs (fiege-fiat-shamir)** for checking the authenticity of the server. Coded a client to connect to the professor’s server and successfully defended from eavesdropping and hacking.