**DIKSHITHA SAI BADIGA**

[dikshitha.softengg@gmail.com | (813)](mailto:dikshitha.softengg@gmail.com%20|%20%20(813)) 540-2364 |  [https://www.linkedin.com/in/dikshitha-s |](%20https://www.linkedin.com/in/dikshitha-s%20|) Tampa, FL

**CAPABILITIES**

• Having **10** years of IT Industry experience with specialization in analyzing complex data sets, providing actionable insight, driving business growth.

• 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.

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.

• 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.

• Proficient in leading Power BI platform development, administration, and training.

• Strong technical design and architecture skills, ensuring alignment with company standards.

• Extensive experience in developing **Power BI** reports and data models that meet business requirements.

• Solid understanding of business needs and usage patterns, contributing to **BI** strategy and roadmap.

• Proven ability to implement governance, access controls, and data segmentation in **Power BI**.

• Skilled in automating workflows and business processes using **Power Automate** and **Power Apps**.

• 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 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.

• Experience with Data Analytics, Data Reporting, Ad hoc Reporting, Graphs, Scales, PivotTables and OLAP reporting.

• Independently handle **Hadoop** administration in local and on cloud in Linux environment.

• Extracting and modelling datasets from a variety 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.

• Expert in data flow between primary DB and various reporting tools, Expert in finding Trends and Patterns within

 • 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.

**TECHNICAL SKILLS & CERTIFICATIONS**

Analytical Tools: Azure Databricks, SAS, Tableau, Google Analytics, Alteryx, PowerBI, Azure ML Studio, Splunk

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

**EXPERIENCE**

**Data Scientist**| **JP Morgan Chase,** Tampa, FL Feb 2020 – Present

• Developed and implemented machine learning models to **assess credit risk**, resulting in a **10%** reduction in default rates and a **15%** increase in approval rates.

• Spearheaded **Customer segmentation** projects, resulting in a **25%** increase in cross-selling opportunities and a 10% growth in customer lifetime value.

• Developed ML model on Churn prediction of loyalty customers based on their usage patterns, allowing for targeted retention campaigns in Spark/PySpark (Azure Databricks).

• Conducted ongoing monitoring of credit portfolios, identifying emerging risks and recommending proactive strategies to mitigate potential losses.

• Built ML model for **Customer Segmentation** of the bank's customer base, enabling personalized product recommendations and targeted marketing campaigns.

• Collaborated closely with banking product managers and relationship managers to implement personalized cross-selling strategies.

• Conducted in-depth analysis of banking customer data to identify patterns and trends in customer behavior.

• Worked with the retail banking team to implement personalized marketing strategies, including tailored **mortgage offers** and **investment recommendations.**

• Monitored the performance of personalization efforts and provided recommendations for continuous improvement.

• Developed data-driven reports for senior management to track **personalization** initiatives' impact on the bank's bottom line.

• Developed and owned reporting on Loyalty customers acquisition, engagement, and retention metrics to track performance and marketing efficacy using **Alteryx** and **SAS** **EG**.

• 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.**

• Utilized **Alteryx** to design and automate complex data workflows, resulting in a 30% reduction in data processing time.

• Assisted senior data scientists in performing data extraction, transformation, and loading (ETL) tasks using **Alteryx**.

• Conducted a market segmentation analysis using **Alteryx** and **Tableau** to identify target customer segments.

• Developed a sales forecasting model using **Python** and **Alteryx** to predict monthly sales trends accurately.

• Created custom dashboards in **Splunk** to present key performance indicators (KPIs), trends, and anomalies, facilitating data-driven decision-making.

• Integrated **Splunk** with other data sources and tools, enabling a holistic view of data across the organization.

• Integrated external market data with internal sales data through **Alteryx** workflows to enhance the model's accuracy.

• Conducted performance tuning and optimization of **Splunk** queries and data collection processes for efficient data processing.

• Built Data pipeline to ingest the statements data from Data Lake to Snowflake using AWS Lambdas, S3 event notifications.

• Developer ETL pipelines using Spark, **PySpark**, AWS EMR, Athena, and Redshift.

• Used **Splunk's** visualization tools to create graphs, charts, and reports for communicating findings to stakeholders.

• Managed Power BI governance, access, and data connection segmentation.

• Developed Databricks and **PySpark** jobs to join Cassandra tables, ETL workflows to load Data into various Targets databases.

• Developed and implemented automated workflows and business processes using Power Automate and Power Apps.

• Performed **NLP** based tokenization, lemmatization, vectorization and developed NLP models for **Topic Extraction and 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.**

• Loyalty Customer’s Life expectancy and the correlation visualization in **Kibana**. (The data was cleaned and processed in excel and the visualization and correlation were done in Kibana)

• 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 PostgreSQL databases.

• Assisted in conducting **A/B testing** across multiple channels that drive improvements in core menu pricing, execution, and experience.

Data Scientist | **American Express, Dallas, Texas** Mar 2018 – Dec 2019

• 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.

• Ingested data to/from Snowflake from S3 based Data Lake with Airflow, Glue/Catalog.

• 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.

• Created workflows for scheduling and data orchestration **ETL** processes i.e., both batch and incremental data loading.

• 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.

**Data Scientist** | **Riktam Technology Cons. Pvt. Ltd.**|Hyderabad, IndiaOct 2013 – Dec 2017

* Led data collection, preprocessing, and cleaning initiatives, employing **SQL** to maintain data integrity and consistency.
* Orchestrated complex **SQL** queries, extracting mission-critical information from **databases,** and contributing to robust business reporting.
* Designed and developed impactful data visualizations using **Excel** and **PowerPoint**, empowering stakeholders with intuitive insights.
* Collaborated closely with business teams to define and track **KPIs**, presenting reports and **dashboards** to drive data-driven strategies.
* Applied **advanced** **statistical** methods to unravel customer behavior trends, resulting in a **30%** boost in customer retention.

**EDUCATION**

**University of Texas at Arlington**| **Master’s Degree** - Computer and Information Science – Dec 2019

**Jawaharlal Nehru Technological University** | **Bachelor’s Degree** - Computer Science and Engineering – May 2013

Courses: Machine Learning, Artificial Intelligence, Intelligence Data Analysis, Database Management System, Operating Systems, Network Security, Advanced Algorithms, Image Processing, Software Engineering, Software Testing.

**PROJECTS**

**Detection of cancer cells in human blood *(University of Texas at Arlington)***

• Engaged in a collaborative effort with a research team to create an application that utilizes **OpenCV** image processing techniques to automatically quantify tumor cells in image samples.

**Android Malware Detection and Security (*University of Texas at Arlington)***

• Established secure communication with **Diffie-Hellman**, used **RSA** and zero-knowledge proofs for server authentication, and designed a client defense against eavesdropping and hacking.