



**Phanindra Nidamanuri**

Data scientist

Speaker phone with solid fill +1 (913) 558-9495

Email with solid fill [phanindra4211@gmail.com](mailto:phanindra4211@gmail.com) psdnnfnbfbffff\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*p\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**OBJECTIVE**

Experienced Data Scientist with 4 years of expertise in analyzing complex datasets to drive strategic decision-making. Proficient in Python, SQL, machine learning, and data visualization tools like Tableau and Power BI. Seeking to leverage data-driven insights to optimize business performance and deliver innovative solutions.

**EDUCATION**

Masters in Computer Science, University of Central Missouri, Kansas, USA

**PROFILE SUMMARY**

* Skilled Data Scientist with over **4 years** of expertise in **data analysis, predictive modeling, and machine learning algorithms** to drive data-driven decision-making and solve complex business challenges across various industries.
* Expertise in **Python, R, and SQL**, with hands-on experience using advanced tools like **TensorFlow, PyTorch, and Scikit-learn** to develop, fine-tune, and deploy machine learning models in production environments for real-time analytics.
* Experienced in **big data technologies such as Hadoop and Spark**, along with cloud platforms like **AWS, GCP, and Azure**, ensuring efficient processing, storage, and analysis of large-scale structured and unstructured datasets.
* Expertise in **statistical analysis, regression modeling, clustering, and A/B testing**, applying advanced analytics techniques to extract actionable insights, improve business performance, and optimize decision-making processes.
* Adept at working with **SQL databases (MySQL, PostgreSQL) and NoSQL databases (MongoDB, Cassandra)**, ensuring efficient data storage, retrieval, and management for large-scale analytical workflows in production systems.
* Strong background in **natural language processing (NLP), computer vision, and AI-driven solutions**, with experience developing **ETL pipelines** to clean, transform, and structure raw data for accurate and meaningful analysis.
* Proven ability to create compelling data visualizations and dashboards using tools like **Tableau, Power BI, Matplotlib, and Seaborn**, effectively communicating insights to both technical and non-technical stakeholders.
* Experienced in building and automating **CI/CD pipelines** for **machine learning models**, ensuring smooth deployment, continuous integration, and faster model iterations in highly scalable environments.
* Familiar with version control systems such as **Git and GitHub**, streamlining collaboration, code tracking, and version management across cross-functional data science teams.
* Collaborative team player with experience working in **Agile environments**, actively engaging in sprint planning, code reviews, and mentoring junior data scientists to foster a culture of innovation and learning.
* Enthusiastic about leveraging **AI, deep learning,** and **emerging data science technologies** to build innovative solutions that enhance operational efficiency and optimize business strategies for long-term success.
* Hands-on experience in developing, training, and fine-tuning deep learning models using frameworks like **TensorFlow, Keras, and PyTorch**, enabling accurate predictions and real-time decision-making for complex business applications.
* Proficient in time seriesforecasting, anomaly detection, and recommendation systems, leveraging advanced machine learning techniques to optimize business operations and enhance customer experiences.
* Expertise in deploying scalable machine learning models using **Docker and Kubernetes**, ensuring efficient model serving, versioning, and seamless integration into cloud-based production environments.
* Strong understanding of **reinforcement learning, generative AI,** and **transformer-based models**, applying cutting-edge AI methodologies to solve challenging business problems in diverse domains.
* Experienced in designing and implementing automated feature engineering pipelines using tools like **Featuretools and MLflow**, significantly improving model performance and reducing manual data preprocessing efforts.
* Skilled in graph analytics, social network analysis, and knowledge graphs, utilizing tools like **Neo4j and NetworkX** to uncover hidden patterns and relationships in large-scale interconnected data.
* Passionate about ethical AI, data privacy, and bias mitigation in **machine learning models**, ensuring compliance with industry regulations like **GDPR and HIPAA** while building fair and transparent AI solutions.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Databases:** | Oracle MySQL SQLite NO SQL, RDBMS, SQL Server 2014, HBase 1.2 MongoDB 3.2. Teradata. Netezza Cassandra |
| **Database Tools:** | PUSQL Developer, Toad, SQL Loader, Erwin. |
| **Web Programming:** | Html. CSS, Xml JavaScript. |
| **Programming Languages:** | R. Python. SQL. Scala. UNIX. C. JAVA. Tableau DWH |
| **Bl Tools:** | Data Stage 9.1. 11.3. Tableau Desktop D3.j s |
| **Machine Learning:** | Regression, clustering, SVM, Decision trees, Classification, Recommendation systems. Association Rules. Survival Analysis etc. |
| **Data Visualization:** | Quick view Tableau 9.4/9.2, ggplot2 (R), D3, Zeppelin |
| **Big data Framework:** | HDFS, MapReduce Pig Hive. Sqoop, Oozie Zookeeper, Flume and HBase Amazon EC2, S3 and Red Shift) Spark, Storm, Impala Kafka |
| **Technologies/Tools:** | Azure Machine Learning SPSS, Rattle Cafe Tensor flow, the anon, Torch Kara’s NumPy. |
| **Scheduling Tools:** | Autosys, Control-M |

**WORK EXPERIENCE**

**Garmin**

Olathe, Kansas, USA

**Role: Senior Data Scientist Sep 2024 - Present**

**Description:** Garmin is a leading product-based company specializing in GPS technology and wearable devices for automotive, aviation, marine, outdoor, and fitness markets. I leveraged machine learning, predictive analytics, and big data technologies using Python, R, and Spark to develop forecasting models and anomaly detection systems. Utilizing AWS and GCP cloud platforms, I implemented scalable AI-driven solutions to enhance operational efficiency and customer experience.

**Responsibilities:**

* Developed and deployed **machine learning models** for product recommendation, allocation planning, and performance optimization. Ensured seamless integration into production environments, improving decision-making efficiency.
* Utilized **Azure Machine Learning (AML)** for scalable model training, evaluation, and deployment. Automated machine learning workflows to enhance predictive analytics capabilities and reduce manual intervention.
* Leveraged **Apache Spark** for distributed data processing, efficiently handling large-scale datasets across multiple sources. Optimized data transformation pipelines to improve real-time analytics and operational insights.
* Conducted **data cleaning, feature engineering, and scaling** using **pandas** and **NumPy** in **Python** to enhance model performance. Applied statistical techniques to remove data inconsistencies and improve accuracy.
* Integrated **machine learning algorithms** into production systems using **Spark MLlib**, **R**, and **Python**, optimizing forecasting and anomaly detection. Implemented real-time prediction models to streamline logistics operations.
* Worked with various databases, including **SQL, NoSQL, and XML**, ensuring efficient data extraction, transformation, and validation. Designed optimized queries to support high-performance data retrieval and reporting.
* Applied **dimensionality reduction techniques** like **Principal Component Analysis (PCA)** and **A/B testing** to refine predictive models. Utilized **ROC plots** and **K-fold cross-validation** to improve statistical significance and model reliability.
* Built **data-driven solutions** using **Python (pandas, NumPy, seaborn, SciPy, scikit-learn, NLTK)** to improve internal processes. Assisted in strategic decision-making by providing actionable insights through advanced data analysis.
* Developed **classification and regression models** using **Python (scipy, numpy , pandas)** and evaluated performance with **R**. Conducted **hyperparameter tuning** to improve the accuracy and generalization of predictive models.
* Created **visualizations and dashboards** using **Tableau, R-Shiny, and Power BI** to present analytics findings. Designed interactive reports to provide business leaders with real-time insights for data-driven decision-making.
* Designed and deployed **NLP models**, **Decision Trees**, **Random Forests**, **K-means Clustering**, and **Naïve Bayes** algorithms using **TensorFlow** and **Scikit-learn**. Developed text analytics solutions for sentiment analysis.
* Built **Python-based APIs** to support customer analytics, integrating multiple **machine learning techniques** for user behavior prediction. Improved marketing segmentation by leveraging AI-driven insights.
* Implemented **customer segmentation** using **K-means clustering**, developing end-to-end data analytics solutions. Integrated **custom visualization tools** in **R, Tableau, and Power BI** to enhance business intelligence reporting.

**Environment:** Azure Machine Learning (AML), Apache Spark, Python, R, SQL, NoSQL, scikit-learn, TensorFlow, Tableau, Power BI.

**Seaboard Corporation**

Merriam, Kansas, USA

**Role: Senior Data Scientist Nov 2023 – Aug 2024**

**Description:** Seaboard Corporation is a diversified conglomerate, operating primarily in agribusiness, pork production, commodity trading, and marine transportation. I developed and deployed machine learning models using Python, TensorFlow, optimize predictive analytics, and improve healthcare outcomes. Additionally, I utilized SQL, Apache Spark, and Databricks to process large-scale datasets, ensuring real-time insights.

**Responsibilities:**

* Implemented various machine learning algorithms such as decision trees, regression models, neural networks, deep learning, and NLP to analyze **transportation** data. Leveraged **Python**, **Scikit-learn**, **TensorFlow**, and **Matplotlib**.
* Performed data preprocessing, feature selection, and transformation using **MLlib** in **PySpark**, ensuring high-quality input data for predictive analytics. Applied feature engineering techniques to improve model accuracy and reduce bias.
* Designed and built real-time dashboards using **Tableau** and **Power BI** to visualize **transportation** metrics, operational performance, and logistics efficiency. Created interactive reports to help stakeholders monitor key performance indicators.
* Designed and implemented real-time data ingestion pipelines using **Apache Kafka** and **Azure ML**, handling high-throughput **transportation** data for timely insights. Ensured seamless integration of streaming data sources.
* Used **Apache Spark** and **Azure Databricks** to process and analyze massive **transportation** datasets, enabling advanced machine learning applications. Optimized distributed computing performance for handling large-scale data.
* Conducted clustering techniques like **K-Means** and **Hierarchical Clustering** using **IBM SPSS** to segment vehicle types, routes, or customer behaviours and identify optimization opportunities. Developed predictive models for delivery delays.
* Created scalable big data processing workflows using **Spark**, **Scala**, and **Python** in **Hadoop** environments, optimizing data extraction, transformation, and analysis. Integrated structured and unstructured **transportation** data from multiple sources.
* Built predictive models such as **Random Forest**, **Decision Trees**, and **Support Vector Machines (SVM)** to forecast route efficiency and optimize maintenance schedules. Applied ensemble learning techniques to enhance model robustness.
* Created **ETL** workflows using **SQL**, **Hive**, and **Pig** to extract, transform, and integrate **transportation** data from multiple sources into a structured and usable format. Automated data cleansing and validation processes to ensure accuracy.
* Applied cross-validation techniques, **log loss** function, **ROC curves**, and **AUC scoring** to enhance model accuracy and identify the most relevant **transportation** predictors. Conducted model performance evaluations.
* Analyzed **transportation** data trends using autocorrelations and time series patterns, ensuring early detection of anomalies in route planning and fleet operations. Implemented anomaly detection algorithms to predict equipment failures.
* Utilized **PCA**, **factor analysis**, and statistical significance testing to refine high-dimensional datasets and improve model interpretability. Applied dimensionality reduction techniques to extract the most informative features.
* Retrieved **transportation** records using **Hive** and **SQL** from **Oracle** and **Hadoop** clusters while using **Python (Pandas, NumPy)** for feature engineering and preprocessing. Developed automated scripts to standardize logistics data formats.

**Environment:** Python, Scikit-learn, TensorFlow, PySpark, Tableau, Power BI, Apache Kafka, Azure ML, Apache Spark, Azure Databricks, SQL, Hive, Hadoop, Random Forest, Decision Trees, SVM

**Bank of America**

Bangalore, India

**Role: Data Scientist Sep 2022 – Jul 2023**

**Description:** Bank of America is a leading global financial institution, utilizing data science, predictive modeling, and artificial intelligence to enhance customer experience and manage financial risks. As a Data Scientist, I worked with distributed data platforms like Hadoop, Apache Spark, SQL, and AWS services such as S3 and Redshift, alongside Python, R, and machine learning algorithms to build models for financial predictions and process optimization.

**Responsibilities:**

* Conducted **BI applications** and **database system studies**, coordinating with team members for system design, integration, and application maintenance, utilizing **Informatica PowerCenter** for **SCD mappings**.
* Developed automated retraining pipelines using **Docker** and **Kubernetes** to ensure models were periodically updated, leveraging **Apache Spark** for large-scale data processing across distributed systems.
* Deployed machine learning models using **AWS Lambda** for serverless model inference and implemented **CDC** for maintaining historical data in **slowly changing dimension tables**.
* Developed, implemented, and maintained **conceptual**, **logical**, and **physical data models** to ensure effective data storage and processing. Participated in the entire project lifecycle, including requirements gathering, analysis, design, development, testing, and deployment of machine learning models.
* Involved in **data validation** using **Tableau**, validating results by querying databases and comparing values with data tables.
* Designed and optimized data connections, data extracts, and scheduled background tasks for weekly and monthly dashboard reports on **Tableau Server**.
* Delivered end-to-end **machine learning projects**, from data aggregation and exploration to building and validating predictive models for business impact.
* Performed **data extraction** and queries on various databases, handling both **structured** and **non-structured data**.
* Utilized **Python 2.x/3.x** to parse and structure raw data, conducting **power analysis** to determine sample strength and performing variable analysis for hidden insights in the data.

**Environment:** Informatica PowerCenter, Docker, Kubernetes, Apache Spark, AWS Lambda, Tableau, Python 2.x/3.x, Tableau Server, SQL, ETL.

**Skechers**

Bangalore, India

**Role: Data Analyst Mar 2021 – Aug 2022**

**Description:** Skechers is a leading American footwear and apparel brand known for its comfort-driven lifestyle and performance shoes, operating globally including in India. As a Data Analyst, I leveraged databases, data visualization tools, and spreadsheet software to analyze large datasets, uncover trends, and provide actionable insights for business decisions. I applied statistical tools and forecasting techniques to improve processes and support strategic planning.

**Responsibilities:**

* Integrated data from multiple sources to ensure data accuracy and integrity, using **SQL** and **Python** for data updates and manipulation, and leveraging **Excel** and **Tableau** for analysis and presentation.
* Performed analyses on both **structured** and **unstructured data** to solve complex business problems using **advanced statistical techniques** and **mathematical analyses**.
* Developed and deployed advanced models using techniques like **multivariate regression**, **logistic regression**, **random forests**, **decision trees**, and **clustering**, and used **XGBoost** for predictive modeling.
* Used **Pandas**, **NumPy**, **Seaborn**, and **scikit-learn** in **Python** for developing various machine learning algorithms.
* Built and improved models using **natural language processing (NLP)** and **machine learning** to extract insights from unstructured data, and utilized **Tableau** for exploratory data analysis (**EDA**).
* Applied **predictive analysis** and **statistical modeling** techniques to analyze customer behavior and offer customized product recommendations. Utilized **Azure AutoML** to automate the training and **hyperparameter tuning** of machine learning models, enhancing the efficiency of model deployment.
* Applied **machine learning techniques** to explore new markets and customer segments, resulting in a 5% increase in customer base and a 9% increase in customer portfolio.
* Collaborated with business partners to understand their objectives, developed **predictive models**, performed **statistical analyses**, generated **data reports**, and tracked **performance metrics**

**Environment:** SQL, Python, Excel, Tableau, XGBoost, Pandas, NumPy, Seaborn, scikit-learn, Azure AutoML, Apache Spark.