**Navya Chowdary**

**Phone:** **+1 312 897 6425**

**Email**: vamshimuniga@braintreesolutions.com

**PROFESSIONAL SUMMARY:**

* Accomplished senior Data Scientist with a history of delivering impactful data-driven solutions through advanced machine learning techniques and statistical analysis.
* Strong expertise in supervised and unsupervised machine learning algorithms, deep learning models, and data mining techniques, combined with a deep understanding of business needs and goals.
* Around **10 years** of experience in analytics with a focus on **data infrastructure and machine learning**.
* Hands-on experience in implementing various Language Model (LLM) techniques, including OpenAI's GPT and Google's BERT, for natural language processing tasks. Leveraging BERT's deep contextual understanding and GPT's generative capabilities to enhance text generation, language translation, sentiment analysis, and other NLP applications.
* Extensive knowledge of advanced analytics methodologies, including predictive modeling, time series analysis, natural language processing (**NLP**), and computer vision, ensemble techniques such as Random Forests, XGBoost etc. and Neural Networks like Convolutional Neural Networks (**CNN**), and Recurrent Neural Networks (**RNN**).
* Proficient in programming languages such as **Python, JavaScript, and SQL**, as well as popular data science libraries and frameworks including **TensorFlow, Keras, scikit-learn,** and **PyTorch**.
* Proficient in conditional text generation, where the **AI can generate text** or content based on specific input conditions, making it suitable for chatbots, content recommendation engines, and personalized content generation.
* Demonstrated ability to design and develop end-to-end data science solutions, from data collection and preprocessing to model building and deployment, leveraging techniques such as cross-validation, A/B testing, and model interpretation.
* Proactive problem solver with the ability to identify business challenges and translate them into actionable data science projects, utilizing techniques such as exploratory data analysis (**EDA**) and hypothesis testing.
* Collaborated with cross-functional teams to deploy **generative AI** models in production environments, ensuring scalability, reliability, and real-time generation capabilities.
* Excellent communication skills, both verbal and written, enabling effective collaboration with cross-functional teams and stakeholders, and translating complex technical concepts into clear business insights.
* Strong leadership qualities with experience in mentoring junior data scientists and leading data science projects, fostering a culture of innovation and continuous learning.
* Proficient in utilizing Azure Databricks for distributed computing, enabling parallel processing and efficient resource utilization.
* Proficient in designing and implementing scalable data architectures to support efficient data processing, storage, and retrieval, leveraging technologies such as Apache Spark, distributed databases, and cloud-based services (AWS, GCP).
* Utilized **generative AI** for data augmentation, generating synthetic data to enhance model training, especially in scenarios with limited real-world data.
* Proficient in handling large datasets, distributed computing (Apache Spark), and cloud services (AWS, GCP).
* Applied NLP, computer vision, and reinforcement learning for tasks like sentiment analysis and image classification.
* Well-versed in SDLC, experienced in Waterfall and AGILE methodologies.
* Applied generative AI for creative content generation, including generating art, music compositions, and even generating code snippets or design concepts.
* Transformed business requirements into data models and algorithms for structured/unstructured data.
* Expertise in Pandas, NumPy, Scikit-learn for machine learning models (e.g., Linear Regression, Random Forest, Neural Networks).
* Experience in multilingual NLP, enabling the development of solutions that can process and generate text in multiple languages, facilitating global reach and engagement.
* Utilized LLMs to create multimodal content, generating textual descriptions for images, enhancing videos with natural language captions, and augmenting data synthetically to bolster model robustness, particularly in scenarios with sparse real-world data.
* Proficiently integrated Azure Databricks with Azure services like Data Lake Storage, SQL Data Warehouse, and Machine Learning, enabling end-to-end data solutions. Utilized Databricks for real-time analytics, facilitating agile decision-making. Established seamless integration with version control systems and CI/CD pipelines for robust and automated model deployment.
* Proficient in custom language model development, tailoring models to specific industry needs or domain expertise, allowing for more accurate and context-aware NLP applications.
* Proficient in Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), two key techniques in generative AI, to generate realistic and novel data, including images, text, and even music.
* Skilled in deep learning techniques with TensorFlow, including CNN, RNN, and regularization.
* Proficiency in model validation, optimization, K-fold cross-validation, PCA.
* Experienced in requirements gathering through interviews, surveys, JAD sessions, wireframes, and flowcharts.
* Familiarity with AWS Cloud services (EC2, VPC, RDS, S3), JIRA, SQL, and data warehousing.
* Utilized Tableau, Excel for data analysis and visualization, and RDBMS like Oracle and MS SQL Server.
* Implemented Bagging and Boosting techniques for enhanced model performance.
* Worked with advanced applications: R, R Shiny, SAS, Plotly, ArcGIS, MATLAB, SPSS.

# TECHNICAL SKILLS

|  |  |
| --- | --- |
| Programming Languages | Python, SAS, JavaScript, SQL, COBOL, JCL, REXX, Visual-Basic, HTML |
| Database | MS SQL Server, Oracle, DB2, Postgres, Redis, MongoDB. |
| Machine Learning | Linear (Simple & Multiple) Regression, Generative AI, Logistic Regression, Unsupervised Learning: Clustering, Support Vector Machine, Principal Component Analysis, Decision trees/ Random Forest, Naïve Bayes, K-Means Clustering, K Nearest Neighbors, Hypothesis Testing, XGBoost, Natural Language Processing (NLP), Probabilistic Graph Model, Reinforcement Learning. |
| Deep Learning | Deep Neural Networks, Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), Bayesian Deep Learning. |
| Natural Language Processing | Lexical processing (Tokenization, Lemmatization, Stemming & Canonicalization), Syntactic Processing (POS Tagging, Viterbi Heuristic, Markov Chain and HMM), Semantic Processing (Arity and Reification, WordNet, and Concept Net Databases). |
| Web/API Framework | Flask, FastAPI. |
| Big Data | Spark, Hadoop. |
| Versioning Tools | Git, GitHub, GitLab. |
| CI/CD Tools | Jenkins, VM, Docker, Kubernetes. |
| Cloud Computing | AWS (EC2, S3, Sage Maker), GCP (Vertex AI). |
| Data Visualization | PowerBI, Tableau, Plotly, Dash. |
| Testing Tool | Postman |

**PROFESSIONAL EXPERIENCE:**

**Sr. Data Scientist**

**VMware, TX (Remote), Nov 2021 to Present**

* Developed a customer clustering project, employing ML and statistical modeling to create predictive models and data products for segmentation.
* Developed Python-based estimation models optimizing gross margins for product bundles, enhancing profitability.
* Built tailored sales models for diverse bundled offerings, maximizing sales efficiency.
* Explored Generative Adversarial Networks (GANs) to craft synthetic data, improving customer segmentation accuracy.
* Engineered predictive causal models using annual failure rates and standard costs, enhancing forecasting accuracy.
* Designed and implemented analytics, machine learning models, and visualizations for comprehensive insights.
* Collaborated cross-functionally with sales and marketing teams, addressing critical data inquiries.
* Prototyped and integrated ML algorithms into production systems, catering to varied business requirements.
* Identified potential warehouse sites using GIS tools, focusing on size and transportation feasibility.
* Crafted Tableau reports, enabling effective tracking of marketing strategies and outcomes.
* Applied Variational Autoencoders (VAEs) to innovate new product bundles based on customer preferences and historical data.
* Proactively identified sales and marketing opportunities, optimizing business strategies.
* Explored Recurrent Neural Networks (RNNs) for crafting natural language product descriptions and marketing content.
* Utilized Azure Databricks for seamless data warehousing and integration, optimizing data workflows.
* Produced georeferenced imagery and maps using GIS, enhancing spatial understanding.
* Developed personalized chatbots using Generative AI, enhancing customer support experiences.
* Conducted precise customer segmentation based on behavior and characteristics, improving service quality.
* Employed CI/CD pipelines for automated testing and API-driven data aggregation.
* Leveraged Azure Databricks as the primary analysis platform, driving various data-centric projects.
* Utilized Generative AI for anomaly detection, testing the resilience of fraud detection systems.
* Applied sentiment analysis using NLP for actionable insights and improved customer satisfaction.
* Integrated AI-driven chatbots for real-time customer support, automating responses and boosting engagement.
* Utilized NER to extract key entities from unstructured text data, enriching customer interaction insights.
* Employed LLMs to personalize responses and product recommendations in customer interactions.
* Integrated NLP with computer vision for comprehensive feedback analysis.
* Employed clustering for personalized marketing using historical, demographic, and behavioral data.
* Utilized Principal Component Analysis (PCA) for thorough high-dimensional data analysis.
* Employed cascading models for credit card data, ensuring robust performance.
* Addressed overfitting and underfitting via meticulous hyperparameter tuning.
* Proficient in A/B testing frameworks, clickstream, and time-spent databases using Airflow.
* Documented process activities and assets in One Trust for transparency and compliance.
* Employed Vertex AI's Explainable AI tools for transparent model decision-making.
* Automated model training pipelines with Vertex AI, reducing development efforts.
* Scheduled model retraining using Vertex AI, ensuring models stayed current with new data.
* Collaborated using Vertex AI's AI Platform Notebooks, fostering teamwork and knowledge sharing.
* Enhanced data access and storage with Vertex AI's integration with Google Cloud Services.
* Leveraged Vertex AI for streamlined model deployment and management.
* Managed large datasets for web app usage and surveys, employing structured and unstructured data.
* Employed K-means Clustering for customer segmentation by demographics.
* Utilized Random Forest and Logistic Regression for quantifying user referral likelihood.
* Developed end-to-end systems with R, Tableau, and Power BI for automation and visualization.

**Environment**: MS SQL Server, R/R studio, Python, Spark framework, Redshift, MS Excel, Tableau, T-SQL, ETL, Seaborn, Sci-kit learn, Keras, TensorFlow, Machine learning libraries, NLP, Linux, Flask, Jupyter, Statistical Analysis, RNN, LSTM MS Access, XML, AWS S3, Sage maker, EMR, Airflow, ML lib, Azure Databricks.

**Senior Data Advisor\Data Scientist**

**Charles Schwab, Austin TX, July 2018 – Oct 2021**

* Implemented Agile Methodology to construct an internal application efficiently.
* Utilized Supply Chain Management to scrutinize supplier data, optimizing costs and enhancing quality.
* Deployed supervised algorithms (Logistic Regression, Decision Trees, KNN, Naive Bayes) for Classification.
* Developed machine learning algorithms in Python (Pandas, NumPy, seaborn, SciPy, matplotlib, scikit-learn, NLTK, Spacy).
* Conducted Exploratory Data Analysis for data extraction and interpretation (Classification, splitting, cross-validation, Regression).
* Proficient in data science tools like R, Python, NumPy, Keras, TensorFlow, etc.
* Employed XGBoost in Python for statistical modeling to predict probabilities.
* Established storage and data analysis tools in AWS cloud infrastructure.
* Utilized AWS SageMaker to create, train, optimize, and deploy advanced ML and DL models.
* Employed Grid Search, K Fold for optimal hyperparameters, and variable selection basis.
* Managed AWS EMR spark cluster for preprocessing vast datasets in S3 buckets using spark data frames.
* Evaluated models via Cross Validation, Log loss function, ROC curves, AUC for feature selection, and elastic technologies like Elasticsearch, Kibana.
* Extensive experience with AWS services such as SageMaker, Lex, Lambda, EMR, S3, Redshift, Quick Sight.
* Completed an immersive Data Science program encompassing Data Manipulation & Visualization, Web Scraping, ML, Python, SQL, NoSQL, MongoDB, Hadoop.
* Implemented diverse ML algorithms and statistical models in Python for accuracy assessment.
* Conducted data collection and mining from primary and secondary sources.
* Applied visualization techniques (histogram, bar plot, pie-chart, scatter plot, box plots) for data assessment.
* Explored and analyzed Sales & Marketing data extensively using R.
* Manipulated and aggregated data from various sources using Nexus, Toad, Business Objects, Power BI, Smart View.
* Developed Restful APIs using Python Flask, SQL Alchemy data models, ensuring code quality.
* Contributed to Database and ETL development, optimizing system performance with query enhancements.
* Implemented classification algorithms (Logistic Regression, K-NN, Random Forests) for Customer churn prediction.
* Utilized Tableau to create analytical Dashboards for data analysis and visualization.
* Promoted data-derived insights for strategic and tactical decisions, emphasizing key metrics/KPIs.
* Applied AI/ML algorithms (decision trees, NLP, regression, neural networks) using scikit-learn in Python, MATLAB for Volume identification.
* Formulated intricate SQL queries for OLAP query log and profiler logs to retrieve dimension and measure usage.
* Leveraged Hadoop ecosystem components (Map Reduce, HDFS, HBase, Scala, Oozie, Hive, Sqoop, Pig, Flume) with installation and configuration.
* Implemented Fast API and Node JS to create APIs for specific applications.
* Created Data Quality Scripts in SQL and Hive to ensure successful data load and quality.
* Extracted unstructured data from MongoDB and performed data aggregation.
* Proficient in version control tools like GitHub.
* Maintained various versions of Test Scripts and executed diverse testing strategies.

**Environment:** R, EC2, EMR, Hadoop, S3, HDFS, Spark (Pyspark, MLlib, Spark SQL), Python (Scikit- Learn/SciPy/NumPy/Pandas/NLTK/Matplotlib/Seaborn), Tableau Server, Machine Learning, NLP, Git Hub.

**Data Scientist**

**CVS, Tx, Dec 2016 to June 2018**

* Processed CSV data with Pandas and NumPy, employing data cleaning techniques.
* Extracted data from databases, transferred to HDFS, and utilized Hive for data retrieval.
* Visualized customer health data to determine optimal insurance pricing.
* Analyzed and visualized data using Matplotlib and Seaborn in Python.
* Vertex AI's AutoML features were utilized to automatically select and fine-tune machine learning algorithms, optimizing model performance.
* Applied feature engineering, normalization, and label encoding with Scikit-learn.
* Used Natural Language Processing for text analytics, including stemming, lemmatization, and stop word removal.
* Developed machine learning models (e.g., Logistic Regression, Gradient Boost Decision Tree, Neural Network) using Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn.
* Collaborated with cross-functional teams to integrate generative AI solutions into existing products and services, enhancing user experiences and personalization.
* Employed Azure Databricks for real-time data streaming and processing, enabling actionable insights from streaming data sources.
* Integrated Azure Databricks with Azure Data Lake Storage (ADLS) for seamless data storage and management.
* Implemented Cross-Validation (K-Fold) to optimize models and prevent overfitting.
* Experimented with Ensemble methods (Bagging, Boosting) to enhance model accuracy.
* Created a distributed random forest with PySpark and MLlib.
* Deployed models using Amazon S3 and SageMaker.
* Utilized AWS SageMaker for building, training, and deploying machine learning models.
* Deployed generative models in real-world applications, such as content recommendation systems, image-to-image translation, and artistic style transfer.
* Collaborated with team members to implement version control and code sharing within Azure Databricks using tools like Git integration.
* Conducted Exploratory Data Analysis, Data cleaning, Feature Scaling, and Feature Engineering.
* Evaluated models using Cross Validation, Log loss, ROC curves, and univariate analysis for feature selection.
* Implemented monitoring, alarms, notifications, and logs for Lambda functions.
* Employed various machine learning algorithms (Logistic Regression, Decision Tree, Random Forest, SVM, Ensemble) for fraud detection.
* Utilized Azure Databricks to create scalable and collaborative data science environments, enabling streamlined data analysis workflows.
* Conducted research and experimentation with different generative models and architectures, fine-tuning hyperparameters to achieve desired output quality and diversity.
* Performed unit testing with accuracy, precision, and confusion matrix.
* Utilized recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks for sequence-to-sequence tasks, such as text generation, language translation, and time series forecasting.
* Stayed updated with the latest advancements in generative AI research and incorporated cutting-edge techniques into projects to achieve state-of-the-art results.
* Collaborated with data engineers to optimize data ingestion and transformation pipelines using Azure Databricks, enhancing data preparation efficiency.
* Integrated Azure Databricks with Azure Machine Learning for model deployment and monitoring, ensuring model accuracy and reliability in production.
* Applied deep reinforcement learning techniques for generative AI, including training agents to generate sequences of actions in various domains, such as game playing and robotics.
* Utilized data science techniques for predictive analysis of medical claims.
* Converted operational Excel reports into Power BI reports.
* Conducted A/B testing to enhance UI/UX and KPIs, boosting revenue by 15%.
* Published reports and dashboards using Tableau.
* Built Confusion Matrix and Classification reports to assess algorithm performance.
* Resolved data quality issues with SQL, efficient coding, macros, and stored procedures for successful job runs.

**Environment**: Python, SQL, Excel, NLP, Neural Network, SQL, Linux, Anaconda, Jupyter notebook, Seaborn, Sci-kit learn, Machine learning algorithms, Dimensionality reduction techniques, AWS Sage maker, Tableau, Power BI

**Data Engineer**

**Union Bank, Hyderabad, May 2015 to Nov 2016**

* Created and maintained database objects such as schemas, tables, views, and stored procedures in Snowflake.
* Loaded the tables from DWH to AWS using Snowflake Cloud Data Platform.
* Implemented Change Data Capture technology in Snowflake to load deltas to a Data Warehouse.
* Develop stored procedures/views in Snowflake for loading Dimensions and Facts.
* Used Snowflake big data components like Hadoop and S3 Buckets and AWS Services for redshift.
* Validated data from SQL Server to Snowflake to make sure it has Apple to Apple matches.
* Worked on Snow SQL and Snow pipe for continuous data load.
* Created data sharing between two snowflake accounts.
* Created internal and external stage and transformed data during load.
* Conducted Data Mapping activity with client to establish the transformation rules from source to target.
* Involved in Designing ETL process that allows reusability with high scalability and maintainability.
* Created global context variables to parameterize the connection details.
* Redesigned the Views in snowflake to increase the performance.
* Created Mappings to populate the data into dimensions and fact tables.
* Implemented Change Data Capture technology in Snowflake to load into Data Warehouse.
* Develop stored procedures/views in Snowflake and used for loading Dimensions and Facts.
* Design, develop, test, implement and support Data Warehousing ETL using Snowflake.
* Good knowledge of RDBMS topics, ability to write complex SQL, PL/SQL.

E**nvironment:** Snowflake, SQL, MS Excel, Visio, MySQL, MongoDB, SQL Server database systems, HDFS, NoSQL, AWS S3 bucket, Spark, Big Data technologies

**Data Analyst**

**UnitedHealth Group, Hyderabad, Nov 2013- Apr 2015**

* Collaborated with the Engineer team to design and maintain MySQL databases for storing and retrieving customer review data.
* Generated DDL and created the tables and views in the corresponding architectural layers.
* Developed intricate SQL queries to access necessary test data, facilitating the creation of comprehensive test cases for thorough validation.
* Worked with Business Analyst’s and Product Managers to frame a problem, both mathematically and within the business context.
* Implemented indexing strategies and query optimizations to enhance MySQL database performance, ensuring efficient data retrieval for customer review analysis.
* Strong Knowledge in object-oriented analysis and design.
* Developed complex SQL queries to find test data and the data needed by the test cases.
* Partnered with Solution Architect to enable appropriate data flow.
* Operated analytical algorithms on HDFS data using map Reduce programs.
* Converted several Informatica workflows into Hadoop, Spark, and Scala.
* Created views in Tableau Desktop that were published to internal team for review and further data analysis and customization using filters and actions.
* Created reports for users in Tableau by connecting to various data sources (MS SQL Server, Oracle, MS Excel, Netezza, .CSV).
* Experience in data analyzing, data cleansing, data manipulating and producing meaningful insights using SQL, MS- Excel, and Tableau.

**Environment**: SQL Database, Data Warehousing, Python/R, Snowflake, Redshift, Data Visualization- SAS/Tableau, Power BI, SQL, PowerShell, Git, and GitHub, MS-Excel.