**Harshitha D**

**harshitha1991d@gmail.com | +19135869599 [|](https://www.linkedin.com/in/harshitha-d-054318313)** [linkedin.com/in/harshitha-d-054318313](https://www.linkedin.com/in/harshitha-d-054318313)

**PROFESSIONAL SUMMARY**

• Accomplished Data Scientist with 10 years of experience leveraging SQL, Python, machine learning, and

data visualization tools like Tableau and Power BI.

• Expertise in developing predictive models, conducting statistical analysis, and delivering actionable insights

to optimize business processes and enhance customer experiences.

• Skilled in data manipulation, preparation, normalization, and predictive modeling using Python and R,

enhancing efficiency and accuracy through advanced data processing techniques.

• Led projects on customer segmentation, pricing optimization, and product recommendation systems using

machine learning and statistical modeling.

• Collaborated cross-functionally to frame critical data questions, prototype ML/DL algorithms, and integrate

them into production systems.

• Managed and analyzed large datasets using platforms like Amazon Redshift, Google BigQuery, and Apache

Spark, applying advanced data cleaning techniques for data quality.

• Designed Python modeling APIs for customer analytics, integrating regression, classification, clustering, and

NLP for user behavior prediction and marketing segmentation.

• Presented insightful dashboards and reports using Power BI, Tableau, and other tools for data-driven

decision-making and strategic planning.

• Experienced in agile methodologies, sprint planning, and cross-functional collaboration, with strong project

management skills.

• Proficient in AWS, Google Cloud Platform (GCP), and Microsoft Azure for deploying and managing cloud-

based data solutions.

• Well-versed in big data technologies like Hadoop, Spark, Kafka, and HBase for optimizing data pipelines and

ETL tasks.

**TECHNICAL SKILLS**

**Business Intelligence Tools**

**Operating Systems**

**Databases**

Tableau (Desktop, Server, Online, Prep), Power BI (Desktop, Online), SSRS, Splunk

Windows 10/8/7, LINUX/UNIX, macOS

MS SQL Server, Oracle, AWS S3, Snowflake, Hadoop

MS Office, MS Project, JIRA, MS Excel, Apache Airflow

**Tools**

**ETL Tools**

SSIS, Power BI, Snowflake, AWS Glue

Python, R, Java, PL/SQL, SQL, C, C++

**Programming Languages**

**Machine Learning and Data**

**Science**

Vertex AI, scikit-learn, NLTK, TensorFlow, GPT-3

Apache Spark, Spark ML, Mllib, Scala, HBase, Kafka, Spark Streaming

Matplotlib, Seaborn, Tableau, Power BI

**Big Data Technologies**

**Visualization and Analytics**

**Cloud Platforms**

Google Cloud Platform (GCP), AWS (Redshift, EC2)

**WORK EXPERIENCE**

**Lululemon, Remote, USA**

**Data Scientist**

**August 2022 - Present**

**Responsibilities:**



•

Conducted ad hoc and operational reporting, manipulating data, and producing routine metrics and

dashboards for management.

• Created parameters, action filters, and calculated sets for preparing dashboards and worksheets in

Tableau.

• Applied advanced data science and AI techniques such as machine learning (ML) and natural language

processing (NLP) to analyze large datasets using Vertex AI, leading to a 43% improvement in predictive

analytics accuracy.

• Utilized machine learning algorithms, including decision trees, regression models, neural networks, and

clustering, using scikit-learn and Python libraries.

• Delivered successful NLP projects, including developing a chatbot for customer support and handling

natural language queries.

• Employed generative AI (GAI) and discriminative machine learning algorithms for customer

segmentation and predictive modeling using Vertex AI.

• Built and maintained data pipelines in Google Cloud Platform (GCP) with Apache Airflow for ETL tasks.

• Developed big data analytics and machine learning applications using Apache Spark with Python, Spark

ML, and Mllib.

• Utilized various data formats such as JSON and XML and implemented machine learning algorithms in

Python.

• Designed and implemented end-to-end systems for data analytics and automation, integrating Tableau

and Power BI.

• Deployed GPT-based NLP models to automate document analysis.

• Developed customized solutions for data visualization using Tableau and Python packages.

**Technologies Used**

SQL, Excel, Tableau, Vertex AI, scikit-learn (decision trees, regression models, neural networks, clustering), NLTK,

TensorFlow, Google Cloud Platform (GCP), Apache Airflow, Apache Spark, Spark ML, Mllib, GPT-3, (Matplotlib, Seaborn).

**Amazon, Austin, TX `**

**Jr. Data Scientist**

**September 2020 – July 2022**

**Responsibilities:**

•

Spearheaded the analysis and preparation of extensive datasets, leveraging historical models to

identify patterns and collaborating with Senior Data Scientists for comprehensive data understanding.

• Conducted data manipulation, preparation, normalization, and predictive modeling using Python and R

to enhance efficiency and accuracy.

• Led a project on customer segmentation through machine learning and statistical modeling,

developing predictive models and data products to support segmentation strategies, increasing

customer targeting accuracy by 35%.

• Developed a pricing model for product and service bundles to optimize gross margins and built price

elasticity models for bundled offerings.

• Under the supervision of Senior Data Scientists, performed data transformation methods including re-

scaling and normalization of variables to enhance model accuracy.

• Designed and developed analytics, machine learning models, and visualizations from prototyping to

production deployment, utilizing Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, MLLib, R, and

various machine learning methods.

• Collaborated with sales and marketing teams to frame and answer critical data questions, prototyping

ML/DL algorithms, and integrating them into production systems for business needs.



• Managed multiple datasets containing structured and unstructured data about web application usage

and online customer surveys, utilizing Amazon Redshift and applying data cleaning processes including

backward-forward filling methods for handling missing values.

• Designed, built, and deployed Python modeling APIs for customer analytics, integrating multiple

machine learning techniques and SDKs for user behavior predictions and supporting marketing

segmentation programs, enhancing customer engagement by 25%.

• Conducted customer segmentation based on demographics using K-means Clustering and explored

regression and ensemble models for forecasting.

• Presented insightful dashboards to Higher Management using Power BI and employed classification

techniques such as Random Forest and Logistic Regression for user referral likelihood quantification.

• Designed and implemented end-to-end systems for Data Analytics and Automation, integrating custom

visualization tools using Power BI, collaborating with project managers and business owners to design

necessary reports.

**Technologies Used**

Python, R, Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, SDK Amazon Redshift, TensorFlow, NLTK, Power BI,

Matplotlib, Seaborn

**Walgreens,IL**

**January 2018 – August 2020**

**Jr. Data Scientist**

**Responsibilities:**

• Developed predictive models utilizing statistical techniques like Bayesian Hidden Markov Models

(HMM) and a variety of machine learning classification models such as XGBoost, Support Vector

Machines (SVM), and Random Forests, leveraging R and Python packages.

• Applied Python-based data manipulation and visualization tools like Pandas, Matplotlib, and Seaborn

to cleanse corrupted data, ensuring accuracy before generating requested business reports.

•

Engineered parsing algorithms to efficiently cleanse and distribute extensive datasets, regularly

handling millions of records.

• Engineered personalized product recommendation systems employing machine learning algorithms

like Collaborative Filtering and Gradient Boosting Trees to cater to existing customer needs and attract

new customers.

• Demonstrated proficiency in implementing Latent Dirichlet Allocation (LDA), Naive Bayes, Decision

Trees, Random Forests, Linear and Logistic Regression, Support Vector Machines (SVM), and Clustering

techniques.

• Demonstrated expertise in developing analytics and statistical models tailored to organizational

requirements, offering cost-effective and efficient alternatives when necessary.

• Extensive experience utilizing Python libraries including Scikit-learn, Pandas, and NumPy.

• Proficient in Normalization and De-Normalization techniques to optimize performance in both

relational and dimensional database environments.

• Conducted data analysis utilizing regression analysis, data cleaning, Excel V-Lookups, histograms, and

TOAD client, providing actionable insights and solutions for investors.

•

Developed scripts to ensure seamless data access, manipulation, and reporting functions using Python

programming language.

• Identified, analyzed, and interpreted trends and patterns within complex datasets utilizing data mining

tools.

• Leveraged Natural Language Processing (NLP) techniques to gauge customer satisfaction and enhance

overall customer experience.



• Ensured all data handling and processing adhered to HIPAA regulations and maintained patient

confidentiality and data security standards.

**Technologies Used**

Python (NumPy, Pandas, PySpark, Scikit-learn, MatplotLib, NLTK), T-SQL, MS SQL Server, HIPAA, Data Lineage,

XML, R Studio, Spyder, MATLAB, ETL, Machine Learning, Shiny, Oracle, Teradata, Java, Tableau.

**AGS Health, Hyd, India**

**Sr.Data Analyst**

**February 2016 – November 2017**

**Responsibilities:**

• Collaborated with subject matter experts and IT teams to define business requirements and transform

them into functional and nonfunctional requirements.

• Developed healthcare analytics and reporting system, utilizing SQL, Python, Tableau, R programming,

SSIS, and AWS cloud services.

•

Conducted analysis on Medicare claims data and beneficiary data using queries developed with the

Medicare Integrated Data Repository.

• Developed quantitative models using Excel and SQL to analyze Compliance and Regulatory

requirements for various financial products.

• Designed and implemented ETL processes using SSIS for data extraction, transformation, and loading,

ensuring data accuracy and consistency.

• Enforced business rules by creating constraints in both production and relational data warehouse

databases.

• Utilized Python and R programming languages for advanced analytics, statistical modeling, and

machine learning algorithms.

• Applied Python libraries such as Pandas, NumPy, and scikit-learn for data manipulation, numerical

computations, and implementing machine learning algorithms.

• Leveraged R programming for statistical analysis, data visualization, and predictive modeling.

• Built interactive dashboards and reports using Tableau, enabling stakeholders to explore and visualize

healthcare data effectively.

• Employed AWS cloud services for hosting the application, data warehouse, and associated

components.

• Developed automated reports and scheduled updates using Tableau's publishing and scheduling

features.

**Technologies Used**

SSIS, AWS Redshift, Tableau 2020, MySQL, Python 3.9, R Studio, AWS EC2

**Cognizant,Hyd,India**

**Data Analyst**

**July 2014 – January 2016**



**Responsibilities:**

• Extracted financial data from various sources such as databases, APIs, and flat files using AWS services

like AWS Glue for data extraction and transformation.

• Utilized SQL queries to extract, filter, and aggregate the extracted financial data from relational

databases.

• Implemented data integration workflows using SSIS to automate the extraction, transformation, and

loading (ETL) processes.

• Cleaned, transformed, and aggregated the extracted data using Python libraries like Pandas to ensure

data accuracy and consistency.

• Implemented dynamic filters, slicers, and drill-down functionalities to enable users to explore and

analyze financial data at different levels of granularity.

• Utilized Python libraries such as Pandas, NumPy, and scikit-learn to develop forecasting models for

predicting future financial trends.

• Applied time series analysis techniques and machine learning algorithms to analyze historical financial

data and generate forecasts.

• Created advanced visualizations using Matplotlib to present complex financial data in a clear and

insightful manner.

• Developed custom plots, charts, and graphs to visualize trends, patterns, and anomalies in the financial

data.

• Leveraged AWS services such as Amazon S3 for data storage, AWS Glue for data integration, AWS

Lambda for serverless computing, and Amazon EC2 for hosting applications.

**Technologies Used:**

AWS, Power BI, Python (Pandas, NumPy, scikit-learn), Matplotlib, SQL, SSIS, Hadoop (HDFS, MapReduce),

Scrum

