**Name: Anjaneyulu**

**OVERALL SUMMARY:**

* Over 8+ years of professional experience as a highly skilled data scientist, Machine Learning, Generative Artificial Intelligence (Generative AI), Software Development, leveraging advanced technologies and Google Analytics to extract valuable insights from complex datasets.
* Expertise in Natural Language Processing / NLP, Generative Artificial Intelligence (generative AI), Deep Learning, Building Data Pipelines, Data Visualization, and Predictive Modeling.
* Proficient in Python for extracting insights and driving business value from complex datasets and Proven track record of using data to solve complex problems and drive innovation.
* Experienced in data cleaning and wrangling, statistical analysis and modeling, machine learning, and data visualization across various industries.
* Developed and deployed machine learning algorithms to solve complex business problems, including regression analysis, classification, clustering, and deep learning as a lead.
* Proficient in Machine learning algorithms like Linear Regression, Ridge, Lasso, Elastic Net Regression, Decision Tree, Random Forests and more Deep Learning algorithms like ANN, CNN, RNN, Ensemble methods like Bagging, Boosting, Stacking.
* Proficient in Python-based machine learning, Deep Learning, Generative Artificial Intelligence (Generative AI) and NLP libraries such as Scikit-Learn, TensorFlow, Theano, Spark Mlib, Keras, PyTorch, TensorRT, cuDNN, NTLK and OpenCV.
* Developed and deployed various LLM (Large Language Models) for Investment Assistance , Customer interaction and identify and mitigate security threats.
* Developed models for applications such as predictive maintenance, recommendation systems, and fraud detection.
* Created an On-Time Delivery dashboard for Supply Chain Management and Vendor Management using Snowflake, Python, and Power BI to improve process flow and showcase vendor performance.
* Built a Last Time Buy Dashboard for Network cards, providing projections for future needs under multiple scenarios using R, Snowflake, and Power BI.
* Designed and developed a business intelligence data visualization dashboard and ETL operations using Tableau Desktop, Power BI, Alteryx, and Quick View, enabling executive management to view past, current, and forecast data.
* Proficient in SQL databases are MySQL, PostgreSQL, MSSQL and NoSQL databases are MongoDB, DynamoDB, Apache Cassandra, HBase and Graph databases are Neo4j, Neptune.
* Designed and developed various analytical reports by blending data from multiple sources in Tableau, utilizing Alteryx for data blending and Power BI with EDL.
* Skilled in data analysis, including data mapping, data cleansing and processing, data conversion programming, and troubleshooting complex data problems.
* Optimized the random forest algorithm, achieving an accuracy of 79% and reducing costs by 8%.
* Involved in the end-to-end process, from data cleaning to deploying code into production and building with REST API services, MLFlow and TFServing.
* Experienced with managing, and deploying large distributed systems like Spark, DASK & H20, and heterogenous platform components and hands-on experience with Apache MXNet, HPC – Nvidia, GPU, CUDA.
* Good team player and quick learner; highly self-motivated person with good communication and interpersonal skills.
* Experience in monitoring and performance analysis of Machine Learning platforms using tools like Grafana, Zabbix and MLops.
* Quick learner in any new business industries or software environment to deliver the best solutions adapted to new requirements and challenges.
* Knowledge and experience in Terraform, Jira, GitHub/Git version control tools, Docker, and Kubernetes with Seldon core.

**TECHNICALSKILLS**

|  |  |
| --- | --- |
| Category | Skills |
| Programming Languages | C++, Python, PowerShell, SQL, T-SQL, R |
| Development Tools | Visual Studio, Git |
| Big Data Technologies | Apache Spark, Hadoop, HBase, HDFS, MapReduce, Apache Hadoop Sqoop |
| Data Science Libraries | Keras, NumPy, OpenCV, Pandas, PySpark, TensorFlow, Scikit-learn, PyTorch, Matplotlib, cuDNN, Apache MXNet, Django |
| Scripting | Shell Scripting |
| Database Technologies | DB2, Informatica, MongoDB, MySQL, Oracle, SQL Server, Sybase,Neptune, Neo4j. |
| Artificial Intelligence | Generative AI, Machine Learning, Deep Learning, Natural Language Processing, Dimensionality Reduction, Logistic Regression, Principal Component Analysis, Random Forests, Supervised Learning, Support Vector Machine (SVM), KNN, ANN, CNN, RNN, Decision Trees, K means, DBSCAN, Reinforcement. |
| Business Intelligence Tools | Power BI, Tableau Desktop, Tableau Server, Microsoft SSRS, SAS |
| Cloud Technologies | Amazon Web Services (AWS), Amazon Elastic Compute Cloud (EC2), AWS S3, Amazon Elastic MapReduce, AWS Glue, Amazon SageMaker, AWS Step Functions, Amazon Comprehend, Amazon Rekognition, Amazon Forecast, Amazon Redshift, Azure Cloud Services, Azure Cognitive Services, Azure Data Bricks, Azure Data Factory, Azure Machine Learning, Snowflake, Azure Data Lake. |
| Agile Methodology | Agile, Scrum |
| Data Quality and Analysis | Data Governance, Data Management, Data Mining, Data Pipelines, Data Profiling, Data Visualization, Data Analysis, Data Cleaning. |
| Web Services | WEB Services |
| Business Planning and Analysis | Business Operations, Business Plans, Sales Forecasting, Market Trends, Business Planning, Statistical Modelling, Marketing Analysis, Customer Engagement, Segmentation, Google Analytics, Abode Analytics, Web Analytics, Finance, Budgets, Pricing, Credit Risk, Risk Management, Business Requirements, Training Programs, Translate |
| User Experience Design | User Experience Design (UX), Visualization |
| Computer Vision | Computer Vision, Sensors, IOT, CUDA, GPU |

**PROJECT EXPEREICNE:**

**Quotient, Mountain view, CA Jan 2022 -Present**

**Lead Data Scientist / Artificial Intelligence Engineer**

**Roles& Responsibilities:**

* Conducted end-to-end data science projects, from data exploration and preprocessing to model development and deployment, using tools such as Python, R, and Azure Machine Learning.
* Worked with large, complex data sets, including structured and unstructured data, to extract meaningful insights and develop predictive models for point of sales (POS).
* Developed and implemented machine learning models, such as deep learning, decision trees, and random forests, to analyze and predict sales patterns and customer behavior at various point of sales (POS) locations and improve customer experiences using MLFlow.
* Develops predictive models using statistical techniques Utilizes regression analysis to understand sales Data in various Customer Transactions using Python and R.
* Analyzed customer behavior data to identify cross-selling opportunities, resulting in a 15% increase in sales and revenue using Python and R.
* Developed and implemented a data-driven solution to reduce fraudulent transactions, resulting in significant cost savings.
* Built a recommendation engine for e-commerce platform, resulting in an 18% increase in customer engagement and satisfaction.
* Integrated Neptune with machine learning algorithms to extract complex patterns and insights.
* Ensured data consistency and reliability through Neptune's ACID transactions.
* Facilitated real-time analytics by implementing Neptune's SPARQL query language capabilities.
* Streamlined data migration processes from relational databases to Neptune graph databases.
* Designed and implemented Chatbot for Customer interaction with Generative Artificial Intelligence / Generative AI technologies, NLP and large language Models (LLM ) tool LangChain.
* Designed and implemented A/B testing and experimentation frameworks to measure the impact of data science solutions on key business metrics.
* Designed and implemented recommender systems which utilized Collaborative filtering techniques to recommend courses for different customers and deployed them to Azure HDInsight.
* Developed and implemented recommendation systems for personalized content and product recommendations, employing collaborative filtering and content-based filtering techniques.
* Collaborated with data engineers and operation team to implement ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.
* Conducts statistical inference to draw conclusions and make predictions. Handles uncertainties and calculates confidence intervals.
* Developed machine learning algorithms for predictive maintenance in the supply chain, optimizing route planning, and improving transportation efficiency using Python and R.
* Employing anomaly detection techniques to identify and mitigate potential risks in the supply chain.
* Utilized data visualization tools such as Power BI to create compelling and actionable visualizations of model outputs, presenting findings to key stakeholders from Sales Data.
* Stayed current with emerging technologies and trends in data science, attending conferences and training programs to expand knowledge and expertise in the field.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL for building (Point of Sales) POS model.
* Identify potential problems and recommend alternative technical solutions.
* Participating in Technical Architecture Documents, Project Design, and Implementation discussions. Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure for large language models (LLM) with LangChain.
* Migration of on-premises data (Oracle/ SQL Server/ DB2/ MongoDB) to Azure Data Lake Store (ADLS) using Azure Data Factory (ADF V1/V2).
* Collaborates with domain experts to understand domain-specific statistical considerations.
* Contributed to the Generative Artificial Intelligence / generative AI and data science community, sharing best practices and insights with other data scientists, Artificial Intelligence Engineer, and collaborating on open-source projects, particularly within the Microsoft ecosystem.

**Environment**: MS SQL server 2014, CUDA, ETL, SSIS, Lang Chain, tableau (desktop 9. /Server 9x), MLFlow, R, Python 3.x(scikit learn/SciPy/NumPy/pandas),Neptune, Neo4j, machine learning (naive bayes, KNN, regressions, random forest, SVM, XGBoost, ensemble), Spark (PySpark, MLlib, Spark SQL), Hadoop 2.x, MapReduce, HDFS, Azure Data Factory, Azure SQL Analytics, Azure Blob Storage, Azure Backup, Azure Files, Azure Data Lake Storage, Azure App Services, Artificial Intelligence ,GPU, Azure Web Apps, Azure Logic Apps, Azure Virtual Machine (VM),NLP, MLOps, cuDNN, TensorFlow, TensorRT.

**GDIT, Rensselaer, NY Dec 2019 – Dec 2021**

**Azure AI Developer / Data Scientist**

**Roles& Responsibilities:**

* Analyzing and interpreting complex financial data, including market trends, investment performance, Web analytics, and risk exposure using Python and R.
* Developing Generative Artificial Intelligence (Generative AI), machine learning and statistical models to support financial decision-making, such as predicting stock prices or identifying investment opportunities using Artificial Intelligence and ML Flow.
* Build and Deployed the Predictive model with Python Django Framework and RESTAPI.
* Developed Google Analytics into a Financial Data System for providing valuable insights and data-driven decision-making capabilities using Python and R.
* Administered Neo4j graph database for relationship-focused data science tasks.
* Developed real-time recommendation engines leveraging Neo4j's native graph processing.
* Integrated Neo4j with existing SQL and NoSQL databases to create hybrid data architectures.
* Spearheaded the use of Neo4j for fraud detection by mapping intricate data relationships.
* Developed the Chatbot for Investment Assistance with Generative Artificial Intelligence (Generative AI) and Natural Language Processing (NLP) using Own Large Language Model (LLM).
* Fine-tune the Generative Artificial Intelligence (Generative AI) model using preprocessed data to enhance its accuracy.
* Used Django's database capabilities, for storing different versions of your financial models and allowed for easy rollbacks and comparisons.
* Collaborating with cross-functional teams, including finance, risk management, and business operations, to develop and deploy data-driven solutions to business challenges.
* Conducting research and analysis to support the development of financial products and services, including creating new investment vehicles or insurance products.
* Incorporates interactivity and user-friendly features in interactive visualizations or dashboards. Creates informative and visually appealing reports and presentations to convey data findings.
* Communicating complex technical concepts and analytical results to non-technical stakeholders.
* Conducting data-driven experiments and A/B tests to validate hypotheses and optimize business outcomes.
* Developed a machine learning model to predict credit risk, reducing default rates by 7% and increasing profitability by 5% using Python and R.
* Engage intensively with cloud aficionados and software developers to meld the Generative Artificial Intelligence (Generative AI) model with cloud-based solutions via APIs.
* Developed and maintained a suite of financial models and dashboards to track and analyze business performance, improving financial decision-making.
* Analyzing supply chain performance metrics and KPIs to identify areas for improvement and recommend actionable insights using google analytics.
* Monitoring key supply chain indicators and providing real-time reports to stakeholders.
* Designed and implemented a data governance program that ensured data quality and integrity, improving regulatory compliance and risk management.
* Conducted research and development in emerging areas of data science, such as explainable Generative Artificial Intelligence (Generative AI), automated machine learning, and causal inference.
* Developed and implemented a fraud detection model using machine learning techniques such as logistic regression, decision trees, and random forests, identifying and preventing fraudulent activity in real-time.
* Collaborates with stakeholders to understand their data visualization needs. Translates complex data analyses into intuitive visual representations for non-technical audiences.
* Addressed intricate issues stemming from Generative Artificial Intelligence (Generative AI) model operations, outputs, and its amalgamation in the software ecosystem.

**Environment:** Hadoop Framework, LLM, HTML, CSS, Java Script, Django, Snowflake, Blob Storage, Spark (PySpark, MLlib, Spark SQL), R, Python 3.X (Scikit Learn/SciPy/NumPy/Pandas/NTLK/Matplotlib/Seaborn / TensorFlow, TensorRT), Tableau Desktop (9.X/10.X), Tableau Server (9.X/10.X), Azure Data Lake, EDL, Machine Learning (Regressions, KNN, SVM, Decision Tree, Random Forest, XGBoost, Lightgbm, Collaborative Filtering, Ensemble), Artificial Intelligence / AI, Natural Language Processing(NLP), OLAP, Teradata, Git 2.X, Agile/Scrum, Amazon Web Services (AWS), Amazon Elastic Compute Cloud (EC2), AWS S3, Amazon Elastic MapReduce, AWS Glue, Amazon Sage Maker, AWS Step Functions, Amazon Comprehend, Amazon Rekognition, Power BI, Amazon Forecast,,cuDNN, Apache MXNet , MLFlow.

**MERCK PHARMA, Branchburg, NJ Nov 2017- Nov 2019**

**AI Engineer/ Data Scientist**

**Roles& Responsibilities:**

* Built and trained natural language processing models for sentiment analysis, text classification, chatbot and named entity recognition, using tools such as LLM (Large Language Model), SpaCy, NLTK, and TensorFlow sing MLFLow.
* Installed and maintained NVIDIA drivers, CUDA toolkit, and associated GPU libraries for optimized performance.
* Developed Generative Artificial Intelligence (generative AI) and computer vision models for image segmentation utilizing frameworks like PyTorch and OpenCV.
* Made significant contributions to the development of Microsoft's Azure Machine Learning platform, designing, and implementing new features and capabilities.
* Piloted the continuous refinement of Generative Artificial Intelligence (generative AI) algorithms by integrating feedback loops with Prompt Flow and Large language Models ( LLM ) with LangChain, enhancing the model's adaptability to evolving user queries.
* Worked with big data technologies like Hadoop and Spark to process and analyze large-scale datasets, leveraging distributed computing for optimized performance.
* Architect and implement ETL/ EDL and data movement solutions using Azure Data Factory, SSIS create and run SSIS Package ADF V2 Azure-SSIS IR.
* Migrate data from traditional database systems to Azure databases. Deploying Azure Resource Manager JSON Templates from PowerShell worked on Azure suite: Azure SQL Database, Azure Data Lake, Azure Data Factory, Azure SQL Data Warehouse, Azure Analysis Service.
* Engage with business users to gather requirements, design visualizations, and provide training to use self-service BI tools.
* Design and implement end-to-end data solutions (storage, integration, processing, visualization) in Azure and Seldon Core for Large Language Models (LLM) with LangChain.
* Collaborated with Microsoft's cybersecurity team to develop and deploy anomaly detection models. These models were designed to identify and mitigate security threats for Large language Models (LLM) with LangChain.
* Conducted research and development in emerging areas of data science, including explainable Generative Artificial Intelligence/ generative AI, automated machine learning / ML, and causal inference.
* Published research papers and presented at conferences on topics related to data science and machine learning, Artificial Intelligence actively contributing to the broader data science community.

**Environment**: azure stack (HDInsight, Azure Machine Learning), Django Large Language Models ( LLM ), LangChain, EDL, ETL, NLP, CUDA, Azure Data Factory, MLFlow,Azure GPU, SQL Analytics, Azure Blob Storage, Azure Backup, Azure Files, Azure Data Lake Storage, Azure App Services, Azure Web Apps, Azure Logic Apps, Azure Virtual Machine (VM), Artificial Intelligence/ AI Azure Cognitive Services(OCR/ICR), Amazon Comprehend, Amazon Rekognition, R scripting, PowerShell scripting & Power BI.

**Data Factz, Hyderabad, India Aug 2016 - Aug 2017**

**Data Analyst**

**Roles& Responsibilities:**

* Developed and maintained financial models and dashboards to track and analyze business performance, utilizing tools such as Excel, SQL, and Tableau.
* Conducted analysis of financial data, including revenue, expenses, and profitability, to identify trends, variances, and opportunities for improvement.
* Built and maintained data pipelines to extract, transform, and load financial data from multiple sources, ensuring data quality and accuracy.
* Collaborated with cross-functional teams, including finance, accounting, and business operations, to develop financial forecasts, budgets, and business plans.
* Conducted ad-hoc financial analyses and prepared reports to support strategic decision-making and business planning.
* Developed and presented data-driven insights to senior management, including C-suite executives, to inform business strategy and performance optimization.
* Conducted data quality audits and implemented processes to improve data accuracy, completeness, and consistency.
* Built predictive models to forecast financial metrics such as sales, revenue, and expenses, using machine learning algorithms and statistical techniques.
* Identified and tracked key performance indicators (KPIs) for the business, developing reporting dashboards to monitor KPI performance.
* Developed and implemented data-driven solutions to address business challenges, such as reducing costs, improving efficiency, and optimizing pricing strategies.
* Prepared test plans to ensure parallel QA and development phases.
* Written and executed test cases, reviewing them with business and development teams.
* Implemented a defect tracking process using JIRA tool, assigning bugs to the development team.

**Environment:** R/R Studio, SAS, SSRS, SSIS, SSAS, Oracle Database 11g, Oracle Bi Tools, Aws, Athena, Tableau, Snowflake, MS Excel, Python, Naive Bayes, SVM, K- Means, ANN, Regression, MS Access, SQL Server Management Studio, SAS E-Miner.

**Xcelvations, Hyderabad, India June 2015- June 2016**

**Jr Data Scientist**

**Roles& Responsibilities:**

* Collect, preprocess, and clean large volumes of automotive data.
* Conduct data analysis to identify trends and patterns in automotive data.
* Established Artificial Intelligence (AI) Model Infrastructure and Performed machine learning modeling to report supervisors about capacity management, resulting in a 10% improvement in workforce transportation efficiency using Artificial Intelligence (AI) / Machine Learning (ML) Flow.
* Created and maintained SQL queries and scripts for data integrity and database maintenance.
* Utilized Excel tools for data validation on various business processes.
* Analyzed data trends, implemented predictive analysis using R, and created visualizations using Power BI for forecasting vehicle performance and maintenance.
* Conducted data manipulation, data cleaning, and addressed missing records to prepare data for analysis, selecting appropriate visualization techniques.
* developed and deployed predictive maintenance models for industrial equipment, utilizing IoT sensors and machine learning algorithms to identify and prevent equipment failures.
* Developed multiple reports and dashboards using a variety of visualization techniques, including dual axes charts, histograms, filled maps, bubble charts, bar charts, line charts, tree maps, box and whisker plots, and stacked bars.
* Experienced in creating ETL packages to extract data from flat files, reformatting the data, and loading it into fact tables.
* Loaded various types of dimension and fact tables into the data warehouse, deployed SSIS packages and solutions into production systems, and generated reports using SSRS.

**Environment:** Programming Languages (PYTHON, R), Power BI, ML Flow Integrated Development Environments (IDEs), Data Manipulation and Analysis Libraries, Artificial Intelligence / AI, Machine Learning Libraries, Frameworks, Data Visualization Tools, Database Management Systems (Microsoft SQL), Cloud Computing Platforms Version Control System, Azure Cognitive Services, Big Data Tools

**EDUCATION:**

* Bachelor of Technology from Jawaharlal Nehru Technological University Hyderabad, India.