

JUNAID SALMAN P

pjunaidsalman@gmail.com | Bracknell | +44 7741333264 | [Linkedin](#)

SUMMARY:

- AI & ML focused Data Scientist with 6+ years of experience delivering data-driven solutions to optimize efficiency, accuracy, and decision-making across diverse industries.
- Specializes in Machine Learning (ML), Artificial Intelligence (AI), Deep Learning (DL), and Natural Language Processing (NLP) applications.
- Proficient in Python and ML libraries: Pandas, NumPy, Scikit-Learn, Statsmodels, Matplotlib, Seaborn, Plotly, ggplot, Pygal.
- Skilled in supervised and unsupervised algorithms: Logistic Regression, Linear Regression, Naïve Bayes, Decision Trees, Random Forest, XGBoost, AdaBoost, KNN, SVM, PCA, K-Means, DBSCAN, Hierarchical Clustering.
- Experienced in deep learning architectures: ANN, CNN, RNN, LSTM using TensorFlow, Keras, and Theano.
- Applied Computer Vision techniques for object detection, image segmentation, and facial recognition.
- Designed and deployed Time Series Forecasting models (ARIMA, SARIMA, Holt-Winters, Prophet, LSTM) for sales, loan demand, and operational predictions.
- Built NLP solutions for topic modelling, sentiment analysis, search optimization, and entity extraction using NLTK, spaCy, and Transformers.
- Created end-to-end AI/ML pipelines for data preprocessing, feature engineering, model training, deployment, and monitoring.
- Strong database expertise with SQL and NoSQL: Oracle, PostgreSQL, MySQL, SQL Server, DB2, Teradata, MongoDB, Cassandra.
- Designed and maintained ETL processes integrating data from databases, APIs, flat files, and CSV sources.
- Experienced in OLAP/OLTP systems, data warehousing, and schema design (Star & Snowflake) using Erwin modelling tool.
- Well-versed in normalization and denormalization techniques for optimizing database performance.
- Expert in Business Intelligence (BI) tools: Tableau, Power BI for building executive dashboards and interactive reports.
- Applied feature selection, dimensionality reduction, and hyperparameter tuning to optimize model performance.
- Experienced with model interpretability tools: SHAP, LIME, and feature importance analysis.
- Implemented MLOps best practices including CI/CD pipelines for scalable AI model deployment and monitoring.
- Managed version control with GitHub (branching, merging, tagging) and containerization using Docker and Docker Compose.
- Familiar with cloud platforms: AWS, Azure, Google Cloud for AI/ML deployments and big data processing.
- Created advanced visualizations to communicate AI insights for both technical and non-technical audiences.
- Conducted Proof of Concepts (POCs) and feasibility studies for AI/ML adoption in business use cases.
- Skilled in handling large datasets through preprocessing, outlier detection, handling missing values, and feature scaling.
- Experienced in integrating API-based AI services for speech-to-text, translation, and vision-based applications.
- Collaborated cross-functionally in Agile and Waterfall environments to deliver production-ready AI solutions.
- Led and mentored data science teams, providing technical guidance, code reviews, and AI project leadership.
- Strong communicator with the ability to simplify complex AI/ML concepts for business decision-makers.

TECHNICAL SKILLS:

Programming Languages: Python, JavaScript, MATLAB, SAS, Spark, SQL, VBA, C++, C, Docker

Python & ML Libraries: NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch, SciPy, Matplotlib, Seaborn, Numba, SpaCy, NLTK, Gensim, Dask, LightGBM, XGBoost, CatBoost, Hugging Face Transformers, OpenCV, Plotly

Machine Learning & AI Techniques: Supervised & Unsupervised Learning, Time Series Forecasting, Natural Language Processing (NER, Sentiment Analysis, Topic Modelling, Text Classification), Generative AI (LLMs, Prompt Engineering), Computer Vision (Image Classification, Object Detection), Support Vector Machines (SVM), Neural Networks (CNN, RNN, LSTM, GANs), Ensemble Methods, Collaborative Filtering, Feature Engineering, Model Optimization & Hyperparameter Tuning

Big Data & Distributed Computing: Apache Spark (PySpark, Spark MLlib, Spark SQL), Hadoop Ecosystem (HDFS, Hive, Pig), Dask, Databricks

Cloud & Data Platforms: Microsoft Azure (Azure ML, Azure Data Lake, Azure Databricks), Amazon AWS (SageMaker, Redshift, Kinesis, EMR, S3), Google Cloud Platform (BigQuery, AI Platform), Snowflake, Heroku, Data Lakes & Data Warehouses, SQL & NoSQL Databases (PostgreSQL, MySQL, MongoDB, Oracle, Teradata)

Data Visualization & BI Tools: Power BI, Tableau Desktop/Server, IBM Cognos 11 Analytics, Plotly, Seaborn, Matplotlib, Jupyter Notebook

Version Control & DevOps: Git, GitHub, GitLab, SVN, Docker, Kubernetes, CI/CD Pipelines

Development Environments (IDE): Jupyter Notebook, Spyder, MATLAB, Visual Studio, VS Code

Methodologies: Agile/SCRUM, CRISP-DM, MLOps, A/B Testing, Model Deployment & Monitoring

PROFESSIONAL EXPERIENCE:

Company: Wunderman Thompson, London.

Sep 2024 - Present

Senior Data Scientist AI

Responsibilities:

- Designed and deployed end-to-end ML, NLP, and Generative AI solutions, reducing processing time by up to 40%.
- Built scalable machine learning pipelines handling datasets of 10M+ records for real-time analytics.
- Optimized model performance through feature engineering and hyperparameter tuning, achieving 15–25% accuracy improvements.
- Developed and maintained Python and SQL scripts for automated data extraction, transformation, and loading (ETL).
- Applied mathematical optimization techniques to improve forecasting accuracy for predictive models.
- Implemented OOP principles, unit testing, and code review practices, reducing production bugs by 30%.
- Leveraged Microsoft Azure Machine Learning and Azure Data Lake for cloud-based model deployment and workflow orchestration.
- Collaborated with cross-functional teams to translate business goals into AI/ML solutions, improving decision-making efficiency.
- Created interactive dashboards and reports in Power BI/Tableau to communicate insights to non-technical stakeholders.
- Established model monitoring and retraining frameworks to ensure ongoing accuracy and relevance.
- Documented processes, models, and code to ensure knowledge transfer and maintainability for future projects.
- Integrated API-based data sources into ML workflows, expanding data coverage by 35% for better model training.
- Conducted A/B testing and model validation to compare algorithms, leading to a 12% uplift in business KPIs.
- Automated repetitive data processing tasks, saving 10+ hours per week in manual effort.
- Implemented LLM-powered sentiment analysis pipelines using BERT and GPT-based models to extract actionable customer insights from multi-channel data sources.
- Applied advanced NLP techniques such as entity recognition, topic modelling, and summarization to enhance content personalization for marketing campaigns.
- Led data governance and compliance initiatives ensuring AI solutions met GDPR, ethical AI, and responsible data usage standards.

Environment: Azure ML, Azure Data Lake, Azure Databricks, Snowflake, Databricks, Hadoop Ecosystem (HDFS, Hive, Pig), Spark (PySpark, Spark MLlib, Spark SQL), Python 3.x (Pandas, NumPy, SciPy, Scikit-learn, NLTK, SpaCy, TensorFlow, PyTorch, Hugging Face Transformers, Matplotlib, Seaborn, Plotly), Machine Learning (Linear/Logistic Regression, KNN, SVM, Decision Trees, Random Forest, XGBoost, LightGBM, Ensemble Methods, Collaborative Filtering), Deep Learning (CNN, RNN, LSTM, GANs), Natural Language Processing (NER, Sentiment Analysis, Text Classification, Topic Modelling), SQL (PostgreSQL, MySQL, SQL Server, Teradata), Git/GitHub/GitLab, Docker, Kubernetes, Tableau Desktop/Server, Power BI.

Data Scientist

Responsibilities:

- Developed Sentiment Analysis model using LLM BERT trained on historical organizational data to assess end-user sentiment, implemented in Python with Google Vertex AI.
- Integrated Guardrails AI with LLM to structure query responses and ensure compliance with business guidelines.
- Performed data collection, cleaning, visualization, and text feature extraction, generating key statistical insights to inform business strategies using Python.
- Applied NLP techniques to classify text datasets, labelling natural language content with relevant predefined categories using Python.
- Managed GitHub repositories, including permissions, branching, and tagging.
- Built an ATM cash demand prediction model using Python and deep learning algorithms, achieving 25% higher accuracy than the legacy method and reducing transportation, logistics, freezing, and insurance costs.
- Created a health insurance claims validation model using ML and DL techniques to detect fraudulent or invalid claims.
- Led and managed research & development efforts for a computer vision team.
- Patented a computer vision and deep learning framework in C++.
- Leveraged big data analytics and domain knowledge to develop multiple business analytics visualizations using Tibco Spotfire, SAP Business Objects, and Crystal Reports.
- Created an ML-based health insurance plan recommendation model to match customers with optimal plans and help organizations reduce losses and liabilities.
- Containerized configuration management tasks using Docker and uploaded built images to local Docker artifact repositories.
- Built One-Class SVM and PCA-based anomaly detection models to identify fraud and irregularities indicative of dishonest behavior.
- Utilized a wide range of Python and R packages: ggplot2, caret, dplyr, RWeka, gmodels, twitter, NLP, reshape2, rjson, plyr, pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-Learn, BeautifulSoup, and development tools like Jupyter Notebook and VS Code.
- Developed integrations with Jupyter Notebook for streamlined experimentation and reporting.
- Built a credit card fraud detection model capable of identifying both fraudulent and legitimate transactions, improving accuracy by 30%.
- Developed predictive models in Python to estimate the probability of attendance for campaigns and events.
- Designed and deployed a Machine Learning CI/CD pipeline on Google Cloud Platform (GCP).
- Conducted sentiment tracking over time using data from multiple sources including the company website, Twitter, Facebook, and Quora.
- Designed and implemented a C++ computer vision library used across all product lines.
- Applied NLP preprocessing techniques — tokenization, stemming, lemmatization, stop word removal, vocabulary phrase matching, and POS tagging — using NLTK and spaCy, converting raw text into structured data.
- Built ARIMA time series forecasting models to analyze sales trends and influencing factors.
- Created sparse datasets using Count Vectorizer, Document-Term Matrix (DTM), and TF-IDF vectorization, assigning unique IDs to words and analysing frequency distributions in the corpus using Python.

Environment: Python (NumPy, Pandas, SciPy, Scikit-learn, Matplotlib, Seaborn, Plotly, ggplot2, gmodels, twitter, NLP, Reshape2, rjson, plyr, NLTK, SpaCy, TF-IDF Vectorizer, Count Vectorizer), C++, R, SQL, PL/SQL, MySQL, Cassandra, Hive, HDFS, Spark (PySpark, Spark SQL, SparkML), AWS (S3, EC2, SageMaker, EMR), Microsoft Azure (Azure ML, Data Lake), Docker, GitHub, Linux, Eclipse, VS Code, Jupyter Notebook, ARIMA, PCA, SVM, BERT, LLMs, Guardrail AI, Hugging Face Transformers, Time Series Analysis, Text Mining, Natural Language Processing (Tokenization, Stemming, Lemmatization, POS Tagging), Computer Vision (OpenCV, Deep Learning Frameworks), TensorFlow, Keras, PyTorch, XGBoost, LightGBM, CatBoost, Anomaly Detection, Ensemble Methods, MLOps (CI/CD pipelines), Data Visualization (Tableau, Power BI), Web Scraping, Social Media Data Mining (Twitter, Facebook, Quora APIs).

Company: L2B Cloud Pvt. Ltd, India.

August 2017 – Jun 2021

Junior Data Scientist

Responsibilities:

- Designed the chatbot for customer queries using NLP.
- Developed and tested Bayesian Estimations of the probability of any action of the internet users (as clicks, conversions) using Real Data.
- Reproduced the working environment with the help of docker to train and run the machine learning model anywhere.
- Developed a highly optimized Sentimental Analysis Model on complaints data in Spark Environment.
- Write samples and guides using Jupyter Notebook.
- Keenly interested in spearheading cross-functional and cross-cultural teams of developers, testers, and data managers for end-to-end delivery lifecycle management.
- Developed and coded Python clear product recommendation process visualization to explain the problem to the business.
- Performed research and developed computer vision solutions to solve law enforcement and biometric problems, such as in visual information retrieval of scars marks and tattoos (SMTs), iris segmentation and iris detection from a distance.
- Good experience in working with various Python Integrated Development Environments like PyCharm, Spyder, Jupyter Notebook, Anaconda.
- Helped in migration and conversion of data from the Sybase database into Oracle database, preparing mapping documents and developing partial SQL scripts as required.
- Instructed the business on how to develop and test website store recommended Advertisements to meet their specific requirements.
- Developed a prediction model to identify high-risk customers and developed a framework to identify anomalies in commercial transactions.
- Generated ad-hoc SQL queries using joins, database connections and transformation rules to fetch data from legacy Oracle and SQL Server database systems.
- Collected data from different cloud platforms like Salesforce Apex, AWS, and Azure.
- Made the model available to end users over the internet using SaaS.
- Worked with all levels of the organization and managed small teams in the project.
- Developed microservices & batches using FASTAPI.
- Created Image Filtering and binary Morphology using NLP.

Environment: Hadoop Framework, Salesforce Apex, Oracle, AWS, and Azure, S3, HDFS, Spark (PySpark, MLib, Spark SQL), Python, Tableau Desktop (9.x/10.x), Tableau Server (9.x/10.x), Machine Learning (Regressions, KNN, SVM, Decision Tree, Random Forest, Collaborative filtering, Ensemble), NLP, Teradata,

EDUCATION:

Visveswaraya Technological University (B.Engr. in Electronics and Communications)

06/21 - 06/21

Royal Holloway University of London (MSc in Artificial Intelligence)

09/23 - 09/24

CERTIFICATIONS: Microsoft Certified Azure Data Engineer [DP-203], Microsoft Certified Azure AI Engineer [AI-102], Wipro Certified Professional in Java & J2ee, Microsoft Certified Azure Data Fundamentals [DP-900], Microsoft Certified Azure AI Fundamentals [AI-900], New Relic Full Stack Observability Practitioner, Business Analytics using Excel, Certified Business Analyst from HRD UAE (CBAP).