**Rahul Kaushik**

• Mumbai, India • +91-826-923-5841 • [kaushikrahul@outlook.in](mailto:kaushikrahul@outlook.in) • [LinkedIn](https://www.linkedin.com/in/kaushikrahul93/) • [GitHub](https://github.com/kaushikrahul1993?tab=repositories) • [Kaggle](https://www.kaggle.com/kaushikrahul)

**PROFESSIONAL SUMMARY**

Results-driven CAE Engineer and Data Scientist with 9 years of experience in a Central Government department, specializing in Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD) simulations, and machine learning applications. Skilled in data analysis, statistical modeling, and developing machine learning solutions. Eager to transition into a Data Scientist role to leverage expertise in Machine Learning, Statistics, and Data Visualization, contributing to impactful, data-driven decision-making in a dynamic organization.

**SKILLS**

* **Statistics:** Scipy, Descriptive Statistics, Inferential Statistics, Hypothesis Testing, ANOVA
* **Data Processing:** Pandas, NumPy, Scikit-Learn, Excel
* **Machine Learning:** Classification, Regression, Clustering, Ensemble Learning, Principal Component Analysis, Hyperparameter Tuning
* **Database Management System Query Language:** Structured Query Language (SQL), MongoDB
* **Data Visualizations:** Tableau, Power BI, Seaborn, Matplotlib
* **Time Series:** Moving Average, ARIMA, SARIMA, ARIMAX, SARIMAX
* **Deep Learning:** Artificial Neural Network (ANN), Convolution Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Encoder-Decoder, Transformer, Bidirectional Encoder Representations from Transformers (BERT), DistilBERT, MobileBERT, TinyBERT, Generative Adversarial Network (GAN), AutoEncoder, Text-To-Text Transfer Transformer (T5), Bidirectional and Auto-Regressive Transformers (BART)
* **Generative AI:** Prompt Engineering, Open AI-GPT, Google-Gemini, Hugging face, Meta-llama, LangChain, Retrieval-Augmented Generation (RAG)
* **Natural Language Processing:** Regex, One Hot Encoding, Bag of Words, N-Grams, TF-IDF, Word2Vec, Glove, Fast text, Sentiment Analysis
* **Deployment:** Flask, Streamlit, Amazon Web Services, Google Cloud, Microsoft Azure
* **MLOps:** Docker, Kubernet, Git, Github, Data Version Control (DVC)
* **Data Engineering:** Hadoop, Apache Spark, HIVE
* **Nuclear Engineering:** Reactor Physics & Engineering, Nuclear Material Transportation, Monte Carlo Particle Transport Simulations, Particle Accelerator Physics & Technology
* **Mechanical Engineering:** Computer aided Engineering (CAE), Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD), Computer Aided Manufacturing (CAM)

**PROFESSIONAL PROJECTS**

**Department of Atomic Energy**, Mumbai, IndiaAug 2015 – Present

**Designation: Scientific Officer**

**Role: CAE Engineer & Data Scientist**

**Title: Development of Control System and Machine Learning Model for Optic Steering in** [**LIGO India**](https://www.ligo-india.in/)**.**

(July 2022 – Present)

* Designed and implemented a control system using Optical Sensors and Electromagnetic Actuators (OSEM) to guide suspension assemblies, achieving positioning accuracy within ±100 micron
* Generated and analyzed over 10,000 data points linking current input to actuator displacement, developing precise control algorithms that improved system responsiveness by 10%.
* Established correlations between actuator displacement and optical sensor output to monitor unexpected movements, resulting in a 20% reduction in positional errors.
* Integrated electromagnetic actuator inputs with optical sensor outputs to improve positional accuracy, achieving a 30% enhancement in stability.
* Developed and optimized predictive regression models (K-Nearest Neighbors, Random Forest, AdaBoost) to predict actuator current input, maintaining constant positioning within ±100 micron; improved model accuracy by 25% compared to initial benchmarks.

**Title: Development of Single/Multi-Stage Suspension Assembly in** [**LIGO India**](https://www.ligo-india.in/)**.** (July 2020 – Present)

* Designed single-stage suspension assemblies with natural frequency modes below 10 Hz; performed modal analysis to identify and optimize fundamental frequencies, ensuring stability and reducing resonance risks.
* Developed CAD models of single-stage suspension assemblies, enabling precise design visualization and alignment with technical specifications.
* Conducted harmonic and random vibration FEA to analyze optic response to ground displacement, achieving an average 1/100th vibration isolation in the 10-100 Hz range, effectively minimizing external interference.

**Title: Quality Prediction of Produced Laser Glass.** (July 2020 – June 2021)

* Collected data on critical parameters (e.g., chamber temperatures, flow rates, mixer speed, Nd doping concentration, glass homogeneity) for the development of a machine learning model to predict phosphate-based laser glass quality.
* Built a Random Forest regression model to predict glass quality using all available features, achieving an initial accuracy improvement of 15% in predicting Nd-doped glass consistency.
* Reduced model complexity with Principal Component Analysis (PCA), effectively reducing dimensions while retaining 95% of the variance in key features.
* Enhanced model accuracy by minimizing Root Mean Squared Error (RMSE), achieving a 20% decrease in RMSE and increasing prediction precision for laser gain medium performance.

**Title: Development of Semi-Continuous Production Facility for Laser Glass Gain Medium.**

(July 2018 – June 2020)

* Designed CAD models of a platinum-based glass production system, enabling precise visualization for production requirements and ensuring alignment with process specifications.
* Performed transient and steady-state thermal analysis to map the system’s temperature profile, identifying critical heat zones and ensuring uniform thermal distribution across the system.
* Calculated electrical power requirements and optimized insulation thickness to maintain the system’s outer surface at room temperature, improving energy efficiency by 100% through effective insulation management.
* Determined optimal flow rate to prevent bubble entrapment in the glass, enhancing glass quality and reducing defects by 25%.

**Title: Anomaly Diagnosis of Steam Turbine** (July 2016 – Jun 2018)

* Gathered condition monitoring and failure data (e.g., turbine speed, steam pressures, temperature, rotor eccentricity, vibration) to develop a machine learning model for early fault detection in nuclear reactor steam turbines.
* Applied Principal Component Analysis (PCA) to reduce model complexity, retaining key features while improving computational efficiency.
* Utilized K-Means clustering to identify optimal fault type clusters, enhancing the system’s ability to diagnose various fault conditions.
* Balanced dataset through SMOTE oversampling, addressing class imbalances and improving model robustness and performance.
* Developed an XGBoost multi-classification model for early fault diagnosis, optimizing accuracy by 25% while preventing overfitting and improving prediction reliability.

**Title: Development of Nuclear Material Transportation Cask (IAEA SSR6 Compliance)**

(August 2015 – June 2017)

* Designed CAD models for nuclear material transportation casks, ensuring structural integrity and compliance with safety standards.
* Conducted dynamic explicit finite element analysis (FEA) to simulate 90 m/s impact tests across various orientations, identifying the most critical orientation for experimental testing and improving safety protocols.
* Performed transient and steady-state thermal FEA to assess cask behaviour at 800°C for 30 minutes, ensuring heat resistance and material stability under extreme conditions.

**PERSONAL/ACADEMIC PROJECTS**

**Title: Customer Segmentation and RFM Analysis for E-Commerce Optimization**

* **Data Import & Cleaning**: Loaded and cleaned a dataset of over **100,000 e-commerce transactions**, removing **100% of duplicates**, handling **5% missing values**, and standardizing **10+ column names** for consistent analysis
* RFM Metrics Calculation: Analyzed customer purchase behavior by calculating 3 key metrics (Recency, Frequency, Monetary) for each customer, covering a base of 50,000+ unique customers to support personalized marketing efforts.
* Customer Segmentation: Segmented customers into loyalty tiers (Platinum, Gold, Silver, Bronze) using RFM quantiles, achieving 20% increase in segmentation accuracy and enhancing targeted engagement initiatives across 4 distinct loyalty levels.
* Visualization & Reporting: Developed a treemap visualization to represent 100% of customer segments, aiding data-driven decision-making. Summarized key metrics by loyalty level, exporting insights that informed strategies for improving customer retention and lifetime value.

**Title: Credit Risk Analysis and Predictive Modeling with Feature Engineering and Imbalanced Data Handling based upon customer’s ‘Account and Payment History’ and ‘Delinquency and Enquiry History’.**

* **Data Cleaning & Transformation:** Handled 5% missing data with KNN imputation, removed 10+ irrelevant columns, and encoded 15+ categorical features to ensure data quality and consistency.
* **Feature Engineering & Selection:** Added 5+ financial behavior features and used VIF to remove 10+ correlated features, reducing multicollinearity and boosting model accuracy.
* **Data Balancing & Modeling**: Balanced classes with SMOTE and RandomOverSampler; built and tuned Random Forest, XGBoost, and ANN models, achieving a **15% improvement in accuracy** through GridSearchCV.
* **Evaluation & Optimization:** Measured model performance with accuracy, precision, recall, F1-score; applied early stopping in ANN for overfitting prevention.
* **Tech Stack:** Utilized Pandas, NumPy, Seaborn, Scikit-learn, and XGBoost for data processing and modeling.

**Title: Machine Learning Model for Employee Attrition Prediction Using Behavioral and Demographic Data.**

* **Exploratory Analysis:** Analyzed **10,000+ employee records** through EDA, identifying key attrition patterns by visualizing behavioral and demographic data with Seaborn and Matplotlib.
* **Data Preprocessing:** Improved model accuracy by processing **20+ categorical and numerical features** using LabelEncoding, StandardScaler, and handling **100% of missing values;** built a custom pipeline for outlier handling, feature scaling, and efficient encoding.
* **Class Balancing & Modeling:** Applied SMOTE to address class imbalance, achieving **balanced distribution across 2 attrition classes;** trained and fine-tuned a Random Forest Classifier with GridSearchCV, resulting in a **15% increase in model accuracy.**
* **Evaluation:** Evaluated model using accuracy and confusion matrix, improving prediction reliability by optimizing the pipeline with imblearn for **enhanced performance stability.**
* **Tech Stack:** Utilized Pandas, Seaborn, Scikit-learn, and imblearn to streamline data processing and model development.

**Title: Robust Predictive Model for Insurance Claim Amount Estimation to Enhance Operational Efficiency and Decision-Making in the Insurance Industry.**

* Data Preprocessing: Preprocessed 100,000+ insurance records by handling 5% missing values, encoding 10+ categorical variables, and scaling numerical features via ColumnTransformer to ensure data quality.
* Feature Selection: Conducted Chi-Square, ANOVA, and Pearson correlation tests, selecting top 15 features impacting claim amount. Used RandomForestRegressor and GradientBoostingRegressor for feature importance analysis, enhancing model interpretability.
* **Dimensionality Reduction: Reduced dimensionality with PCA, capturing 95% variance in 5 principal components, enabling efficient 2D visualization for improved model insights.**
* **Model Building: Developed and evaluated multiple regression models (Linear Regression, AdaBoost, Random Forest, Gradient Boosting), achieving R² score increase of 20% and minimizing mean squared error (MSE) across models.**
* **Model Tuning: Enhanced Gradient Boosting model accuracy with GridSearchCV, achieving a 10% reduction in MSE by optimizing hyperparameters for precise claim amount predictions.**
* Tech Stack: Utilized Pandas, Scikit-learn, and GridSearchCV to streamline data processing, feature selection, and model development.

**Title: Predictive Model for Daily Oil Production Based on Well Attributes and Geological Factors.**

* Data Preprocessing:Processed 10,000+ well records, converting categorical features into numerical values via one-hot encoding, ensuring data consistency.
* Feature Engineering: Created a depth-based feature to enhance predictive accuracy for oil production, contributing to a 10% improvement in model performance.
* Feature Selection: Selected key features with Chi-Square and ANOVA tests, reducing feature space by 30% to focus on top predictors of oil well production.
* Model Development: Built and compared multiple regression models (Linear, Lasso, Ridge, ElasticNet, Random Forest, Gradient Boosting), identifying the most accurate model for oil production estimation.
* Model Evaluation: Assessed model accuracy with R² score and Mean Squared Error (MSE) on training and test sets, achieving a 15% decrease in MSE post-feature selection and tuning.
* Tech Stack: Utilized Pandas, Scikit-learn, and regression modeling techniques to streamline data processing and model development.

**Title: Predictive Modeling of Customer Churn in the Telecom Industry Using Binary Classification**

* Data Preparation: Processed 50,000+ customer records, handling 5% missing data, converting 10+ categorical variables with Label Encoding, and replacing blanks with NaN for data consistency.
* Insight Discovery: Conducted an in-depth EDA, uncovering key customer behavior patterns related to churn through visualizations, enabling targeted feature engineering.
* Significant Feature Selection: Applied Chi-Square, ANOVA, and feature scaling to identify top predictors, reducing feature count by 20% and balancing class distribution with SMOTE for improved accuracy.
* Model Training & Comparison: Developed and tuned Random Forest, XGBoost, Logistic Regression, and LightGBM models, increasing prediction accuracy by 20% post-optimization.
* Performance Assessment: Validated model with ROC-AUC, confusion matrix, and accuracy score, implementing cross-validation and hyperparameter tuning to achieve a 15% boost in precision.
* Tools & Libraries: Utilized Pandas, Scikit-learn, and LightGBM to streamline preprocessing, feature engineering, and model development.

**Title: Web Scraping and Data Analysis for Real-Time COVID-19 Statistics using BeautifulSoup and Pandas**

* Real-Time Data Extraction: Scraped real-time COVID-19 data from multiple sources using BeautifulSoup and Requests, collecting data from 10+ websites for comprehensive analysis.
* Data Cleaning & Transformation: Cleaned and preprocessed raw data by handling missing values and converting 15+ columns into appropriate data types, ensuring consistency for analysis.
* Trend Analysis: Conducted exploratory data analysis (EDA) on global and regional COVID-19 trends using Seaborn and Plotly, revealing key insights into infection rates and recovery patterns.
* Interactive Data Visualization: Developed interactive dashboards and charts with Plotly to visualize key COVID-19 statistics, enabling real-time decision-making.
* Technologies Used: Python (BeautifulSoup, Requests, Pandas, NumPy, Seaborn, Plotly) for efficient data scraping, processing, and visualization.

**Title: AI-based Sentiment Analysis and Recommendation Systems for Cancer Survivors and Caregivers Using Llama 2**

* Text Preprocessing & Cleaning: Processed 50,000+ text entries, cleaning data by handling contractions, removing punctuation, filtering stopwords, and applying lemmatization and tokenization, improving model input quality.
* Feature Engineering with TF-IDF: Transformed raw text into numerical features using TF-IDF, capturing top 20% of important terms to enhance the text classification model's accuracy.
* Model Development with XGBoost: Built and optimized an XGBoost model for sentiment classification, achieving 85% accuracy in predicting mental health-related outcomes, categorizing posts into intensity and predicted status.
* Exploratory Data Analysis (EDA): Conducted EDA on 50,000+ records, visualizing key trends in mental health data using count plots, pie charts, and word clouds to analyze sentiment distribution and intensity.
* Model Evaluation: Evaluated model performance using classification metrics, ensuring 90% accuracy and high precision on both training and test datasets.
* Technologies Used: Utilized Python, XGBoost, TF-IDF, Seaborn, and Pandas for data processing, feature extraction, and model development.

**Title: Time Series Forecasting for Stock Market Prediction Using ARIMA and Prophet**

* Data Preparation: Cleaned and processed 100,000+ stock price records, addressing 2% missing data and ensuring appropriate data types for target variables. Removed duplicates to ensure high-quality data for analysis.
* Decomposition & Trend Analysis: Applied seasonal decomposition to stock price data, extracting trend, seasonality, and residuals to capture the underlying patterns and improve the forecast model.
* Stationarity Testing & Transformation: Conducted ADF tests on stock prices to check for stationarity, and applied differencing to transform the data, resulting in a stable time series for modeling.
* Forecasting Model Development: Developed and trained models using ARIMA and Facebook Prophet, incorporating seasonality and trend components, and achieved 85%+ forecasting accuracy in stock price predictions.
* Prediction & Model Evaluation: Generated 1 to 3-month stock price forecasts and assessed the model’s effectiveness using prediction intervals, resulting in a 15% increase in forecast reliability.
* Tools & Libraries: Leveraged Python (ARIMA, Prophet, Pandas, Matplotlib) for effective data preprocessing, modeling, and result visualization.

**Title: Sentiment Analysis on Twitter Data using NLP and Machine Learning Models**

* Text Cleaning & Preprocessing: Processed 50,000+ tweets, removing stopwords, stemming, and tokenizing for improved data quality.
* Feature Engineering: Extracted features using BoW, TF-IDF, and Word2Vec, boosting model performance by 30%.
* Machine Learning Models: Built Logistic Regression and Random Forest models, achieving 85% accuracy for sentiment prediction using traditional NLP techniques.
* Advanced Deep Learning: Trained LSTM, GRU, and Bidirectional LSTM models, improving sentiment classification accuracy by 10%.
* Transformer-Based Models: Applied pre-trained BERT and GPT models via Hugging Face, reaching 90%+ accuracy for more nuanced sentiment analysis.
* Tools: Python, Hugging Face, Pandas, scikit-learn, TensorFlow for NLP and machine learning model development.

**EDUCATION**

* [**Learnbay**](https://www.learnbay.co/datascience/data-science-and-ai-master-certification-program)**,** Bangalore, India December 2024

Online Master Certificate, Data Science & Artificial Intelligence

* [**Bhabha Atomic Research Centre**](https://www.barc.gov.in/careers/officers.html)**,** Mumbai, India July 2015

Orientation Course, Nuclear Science & Engineering with Distinction.

* [**Shri G.S. Institute of Technology & Science**](https://www.sgsits.ac.in/index.php/departments/mechanical-engineering)**,** Indore, India June 2014

Bachelor of Engineering, Mechanical Engineering with Honors.

* **Indira Gandhi H.S. School,** Morena, India  June 2014

Higher Secondary School Certificate, Mathematics with 98 Marks

* **Govt. Boys H.S. School,** Sheopur, India  June 2014

High School Certificate, Mathematics with 99 Marks

**TRAINING COURSES & CERTIFICATIONS**

* ‘Data Science and AI Certificate for Managers & Leaders’ powered by IBM Skills Network issued by Learnvista. (Nov 2024).
* ‘Deep Learning Fundamentals’ powered by IBM Developer Skills Network issued by Learnvista (Nov 2024).
* ‘Machine Learning with Python’ powered by IBM Developer Skills Network issued by Learnvista (Nov 2024).
* ‘Python 101 for Data Science’ powered by IBM Developer Skills Network issued by Learnvista (Nov 2024).
* ‘Hands on Training on Computational Fluid Dynamics’ organized by ‘Indian Nuclear Society’ (August 2019).
* ‘Basic Radiological Safety & Regulatory Measures for Nuclear Facilities’ organized by ‘BARC Safety Council Secretariat’ (March 2018)
* ‘Application of Numerical Heat Transfer to Industrial Problems’ organized by ‘Indian Nuclear Society’

(May 2017).

* ‘Application of Finite Element Technique in Industrial Problems’ organized by ‘Indian Nuclear Society’

(Feb 2017).

* ‘Design of Nuclear Pressure Vessel & Piping’ organized by ‘Indian Nuclear Society’ (November 2016)
* ‘Executive Excellence Program’ organized by ‘Dale Carnegie & Associates, Inc.’ (April 2015)

**ACCOMPLISHMENTS**

**Title: Group Achieve Award in Excellence in Science, Engineering & Technology Award Scheme**

* ‘Enhanced Mixed Carbide Fuel Fabrication for FBTR & Transportation of Special Nuclear Materials’ (June 2018).
* ‘Development of Transport Package Impact Dampener’ (June 2017).

**Title: Scholarships**

* ‘Central Sector Scheme of Scholarship’ for College and University students (July 2011 – June 2014)