EVA KAUSHIK

(+91) 9667841697 | kaushikeva0026@gmail.com | LinkedIn | GitHub | Medium | Portfolio

PROFESSIONAL SUMMARY

A Data Scientist experienced in delivering business-impacting results through comprehensive data solutions. Proficient in Python, R, and machine learning. Skilled in Generative AI, clustering, and database design, with hands-on MLOps experience on cloud platforms like AWS, Azure, and Databricks. Strong knowledge of Big Data tools (Hadoop, Spark, Hive) and statistical techniques such as regression analysis, Bayesian methods, and time series analysis. Renowned for outstanding analytical skills, capable of translating massive volumes of data into actionable insights. Eager to contribute to cutting-edge Data Science and Statistics research and teaching at a prestigious academic university.

EDUCATION

Guru Gobind Singh Indraprastha University
Bachelors in Technology, IT, GPA 3.66/4.0   *June 2018 - July 2022*
Relevant Coursework: Data Structures, Mathematics I & II, Database Management Systems, Design and Analysis of Algorithms, Artificial Intelligence, Operating System, Software Engineering and Big Data Analytics.
Award/Scholarships: Ranked among the top 3 students in the first year, leading to a branch change from ECE to IT and received an appreciation letter from the college director for outstanding all-round performance.

SKILLS & CERTIFICATIONS

- **Programming Languages:** Python, R, C++, MATLAB, SQL, Java, Shell Scripting
- **Data Science & Machine Learning:** NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, PyTorch, TensorFlow, Keras, OpenCV, MLlib, Data Wrangling, Data Visualization, A/B Testing, Ensemble Learning, Hyperparameter Tuning, Model Evaluation
- **Statistical Analysis:** Regression Analysis, Hypothesis Testing, Bayesian Methods, Probability Theory, Time Series Analysis, Multivariate Statistics, Dimensionality Reduction, Statistical Inference
- **Big Data Technologies:** Hadoop, Spark, Hive, Snowflake, AWS Redshift
- **Cloud Platforms & MLOps:** AWS (S3, EC2, CloudFront, DynamoDB), Azure, Databricks, Docker, Jenkins, Git, GitHub, CI/CD Pipelines
- **Data Management & Visualization:** Tableau, Power BI, Microsoft Excel, Microsoft Access, Google Colab, Jupyter Notebooks
- **Advanced Techniques:** Generative AI Models (StyleGAN, GPT-4, DALL-E), Natural Language Processing (NLP), Computer Vision (CNNs, GANs), Reinforcement Learning, Graph Neural Networks (GNNs)
- **Certifications:** AWS Cloud Practitioner, NVIDIA GSI Technologies (2024), Dell Professional API Design, AWS Generative AI Sales, AWS Technical Accreditation, Dell Associate Administrator, AWS Migration Hub Primer, Statistics for Data Science, Azure Data Scientist (Registered), Databricks Certified GenAI Engineer Associate (Registered), Mathematics for ML
- **Additional Skills:** Data Warehousing, Bayesian Inference, Adaptive Algorithms, Real-time Analytics, Advanced Statistical Methods, Computational Statistics

EXPERIENCE

Data Scientist, Nestle (Contract: DXC India & Poland)   *Feb 2024 - Present*

- Deployed advanced **AWS services (Bedrock, SageMaker) and OpenAI models on Azure**, optimizing autoscaling and resource management. Integrated **Generative AI models(GPT-4, DALL-E, Embeddings v3, GPT-4o mini, AWS Claude 3 Sonnet, others)**, enhancing predictive accuracy and performance. Implemented **RAG templates**, improving efficiency.
- Core Member in production of **Nes-GPT**, implemented nine production-ready models, using machine learning algorithms (deep learning, reinforcement learning) and statistical methods(Bayesian inference, Monte Carlo simulations). Addressed a **$30 million use-case of the SDXL Inpainting Model** [https://nesgpt.genai.nestle.com/]
- Data analysis and visualization with **Big Data tools (Hadoop, Spark, Hive)** and **statistical techniques (regression analysis, hypothesis testing, time series analysis)**. Enhanced autoscaling for higher throughput and cost efficiency by 20%..
- Collaborated with cross-functional teams to design innovative data solutions. Utilized **NLP techniques (transformers, BERT)** and **computer vision (CNNs, GANs) for AI applications**.

Data Scientist/ML Engineer, DXC Technology, Noida, India   *Aug 2022 - Present*
- At DXC, annotation tools were optimized, and GPU cost analysis was conducted for computer vision projects, reducing computational costs. An interactive PyQt5 robot was developed, enhancing UI responsiveness and integrating advanced navigation libraries for improved user interaction.
- Additionally, the implementation of **GenAI-driven Digital Twin technology** was spearheaded, streamlining Eye-Gazing experiments with Python/Azure, which reduced setup time by 50%. AutoTS experiments with **LSTM/Prophet** were conducted, achieving a 15% improvement in predictive accuracy through meticulous hyperparameter tuning and validation using MAE, RMSE, and MAPE metrics.
- A real-time anomaly detection system using **Graph Neural Networks (GNNs)** for financial transactions was engineered, achieving a ~98.3% accuracy rate and significantly reducing fraud detection time. Furthermore, a scalable recommendation engine using Reinforcement Learning was architected, boosting user engagement by 20% and optimizing resource allocation, enhancing overall system efficiency.

Exceptional Performance: With Incisive thinking and consistently out-performing in rigorous interviews, I facilitated my transfer to Analytics Delivery, Data Science at GDN India Capability.

Founding Pioneer, Dexignare   *June 2021 - July 2023*
- Built and used data science algorithms such as machine learning (ML) deep learning (DL), statistical analysis to improve user experience (UX), and interface design. Applied **convolutional neural networks (CNN)** for image classification and object recognition, also **generative adversarial network(GAN)** for synthetic data generation — made the product more user-friendly.
- Learned and implemented high precision models for predictions.Implemented ensample learning. Analyzed data using **Bayesian inference, regression analysis, clustering and principal component analysis (PCA) / t-distributed stochastic** neighboring entities with grid search. Worked with state of the art statistical methodologies such as Hypothesis Testing/ Time series analysis/ Stochastic processes to get actionable insights that led us to set industry benchmarks.
- Learning algorithms and **deep learning networks (dCNNs, dRNNs,dGAN)** to improve UI/UX design solutions. Advanced optimization algorithms : Better operational efficiency

Advisor, Frontforumfocus Kenya, South Africa   *Apr 2021 - Aug 2022*
- Deep learning frameworks (TensorFlow, PyTorch) and advanced algorithms (**XGBoost, Random Forest, Gradient Boosting Machines**) for complex predictive modeling, achieving a **40% increase** in data processing efficiency and a **15% boost** in overall project delivery speed.
- Established robust **CI/CD pipelines** using Jenkins for continuous integration, **Git for version control**, and Selenium for automated testing, improving code quality and reducing deployment failures. **MLOps** to streamline the entire Workflow.
- With data-driven strategies, increasing decision-making and reducing system bugs by 30%. Enhancing data processing speed by **40%.**

PROJECTS

- **HYPERTUNE-X [Github]**
  It is a powerful, lightweight framework designed to manage and scale machine learning experiments efficiently. It supports multiple logging backends, modular configurations, and seamless integration with Docker and Kubernetes. Ideal for individual data scientists and small teams, HyperTune-X simplifies systematic logging and experiment management, boosting productivity by up to 32%.

- **MARKOV-CHAIN-MONTE-CARLO-ALGO[Github]**
  MCMC algorithms for high-dimensional statistical inference and synthetic data generation. Implements adaptive algorithms with Julia, enhancing computational efficiency by 41%. Focuses on rigorous data preprocessing, potential function calculations, and diagnostic measures,

ensuring model robustness and stability with a 92% confidence interval.

- **AUTO-ML** [Github]
  A Python library for automated machine learning on tabular data, featuring automated data cleaning, feature engineering, and hyperparameter optimization. Integrates frameworks like scikit-learn, XGBoost, and LightGBM, enhancing workflow efficiency by 48% for binary classification, regression, and multiclass classification tasks. Achieves up to 30% improvement in model accuracy through advanced hyperparameter tuning.

- **JENKINS-CODEFLOW** [Github]
  A robust CI/CD automation server leveraging Jenkins, integrated with Docker and Kubernetes for scalable deployments. Utilizes YAML configurations for seamless pipeline management. Python scripting and HTML/CSS customization enhance automation, reducing deployment time by 40% and improving workflow efficiency by 35%.

SERVICES AND VOLUNTEERING EXPERIENCE

- **Core Member and Speaker, IEEE Quarter Tech Talk Series,** delivered talks on data science and technology.
- **Active Member, NSS (National Service Scheme),** contributed to various community service projects.
- **Volunteer, Milestone NGO,** supported civil rights and social action initiatives.
- **Social Media Volunteer, IEEE Region 8,** managed social media for event promotion.
- **Organizer, IEEE-USA,** coordinated workshops and seminars.
- **Campus Ambassador, International Model United Nations,** promoted and facilitated model UN activities.
- **Volunteer, MOVE Disaster Relief,** assisted in disaster relief and community support efforts.

RESEARCH EXPERIENCE

Research Assistant
Supervisor: Dr. Kamal Upreti   *Jan 2022 - June 2022*
Publication: [Cardiovascular disease prediction using machine learning](https://doi.org/10.1063/5.0150418), AIP Conference Proceedings, Volume 2587, Issue 1, 2023

Aim: Logistic regression, SVMs, and neural network-based predictive models for cardiovascular disease risk.
Contributions:
Large scale health datasets (collected, pre-processed).
Improved model accuracy using feature extraction stages.
ML algorithms optimized for better performance.
Performed robust statistical analyses and developed visualizations/reports.
Abstract: The study results showed that ML/AI had the potential to detect CVD early and help in preventive healthcare.

Thesis/Capstone Project
Title: Dynamics of EHR in m-healthcare application
Publication: Human-Machine Interface, 295–309, 2023
DOI: [10.1002/9781394200344.ch11](https://doi.org/10.1002/9781394200344.ch11)
Investigated the dynamics of Electronic Health Records (EHR) in mobile healthcare applications, focusing on system integration and data management.

ACCOLADES

- **H2 FY24 DXC** Collaborators Award (2024)
- **DXC Grow Award** for Exemplary Mentorship (2024)
- **DXC Grow Deliver Award**, Strategic Innovator: Technology & Intelligence (2024)
- **Advisory Board Member**, IFERP International Conference (2024)
- **Guest Speaker**, NIT Delhi and other Universities on AI, ML, and Blockchain (2023)
- Selected in **Microsoft for Startups Founders Hub** [MVP-Peable]
- **Jury Member**, "Hack-O-Hiest" Hackathon, IEEE ADGITM (2022)
- **Nominated Binance Build for Bharat Hackathon**, Top 114 Startup Ideations (2022)
- **Best Chairperson**, IEEE ADGITM Student Branch (2021-22)
- **WIE Affinity Group Award**, IEEE WIE AG Delhi Section (2021)
- **Intra-IEEE Project Competition Winner**, Sign-Language Recognition System (2021)
- **Mentorship Award**, Technorax V5.0 Hackathon (2020)
- **1st Rank, Technorax IEEE ADGITM** Hackathon (2019)
- **Merit Scholarship**, Topper, All India Senior Secondary School Examination, CBSE (2016)

TEACHING EXPERIENCE

Data Science Tutor   *July 2022 – Jan 2023*
Superprof (Remote)

- Instructed early-stage graduate students from the University of Maryland in advanced Data Science and statistical methods via the Superprof platform.
- Delivered customized lesson plans on artificial intelligence, statistical modeling, algorithm development, and data visualization.
- Guided the student through complex strategies and projects, enhancing their practical application of theoretical concepts.
- Provided regular assessments and feedback, significantly improving the student's academic performance and confidence.

Relevant Coursework: Machine Learning, Data Mining, Statistical Inference, Data Visualization, Big Data Analytics, Predictive Modeling, and Advanced Statistical Methods.

PUBLICATIONS

[1] Vashisht, C., Kaushik, R., & Kaushik, E. (2024). Chapter 10 semantic-based approach for medical cyber-physical system (MCPS) with biometric authentication for secured privacy. *Digital Transformation in Healthcare 5.0*, 237–266. https://doi.org/10.1515/9783111398549-010 [Publication]

[2] Kushwah, A., Rajora, T., Singh, D., Pandey, S., & Kaushik, E. (2024). Galactic simulation: Visual perception of Anisotropic Dark matter. *Communications in Computer and Information Science*, 25–36. https://doi.org/10.1007/978-3-031-47221-3_3 [Publication]

[3] Kaushik, R., Sachdeva, R., Vashisht, C., & Choudhary, K. K. (2023a). Sensory perception of haptic rendering in surgical simulation. *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. https://doi.org/10.1109/icccnt56998.2023.10306364 [Publication]

[4] Kaushik, R., & Kaushik, E. (2023). Postmortem concentrations: Distributed privacy-preserving Blockchain Authentication Framework in cloud forensics. *Blockchain for Healthcare 4.0*, 161–184. https://doi.org/10.1201/9781003408246-8

[5] Kaushik, E., & Kaushik, R. (2023). Genetically induced biomaterial advances in medical sciences. *Engineering Materials*, 95–123. https://doi.org/10.1007/978-981-99-6698-1_4 [Publication]

[6] Rai, S., Jain, I., Vij, A., & Kaushik, E. (2023). Translational Bioinformatics Ontology in healthcare with cloud computing. *2023 International Conference on Innovations in Engineering and Technology (ICIET)*. https://doi.org/10.1109/iciet57285.2023.10220662 [Publication]

[7] Kaushik, R., & Kaushik, E. (2023a). Postmortem concentrations: Distributed privacy-preserving Blockchain Authentication Framework in cloud forensics. *Blockchain for Healthcare 4.0*, 161–184. https://doi.org/10.1201/9781003408246-8 [Publication]

[8] Dermatological Diagnostics: A Unified Deep Learning Framework for Skin Lesion and Cancer Classification [Accepted in IEEE; Publishing date: Dec 24]

[9] Financial Exclusion: Evolution and Development of Microfinance Statistics in India [Accepted in Springer; Publishing date: Jan 25]

[10] Cutting Edge Hybrid Models for Credit-Card Fraud Detection: Integrating GRFD-KNN, Temporal Analysis, and LSTM Networks [Accepted in Springer; Publishing date: Feb 25]