

# EVA KAUSHIK

(+91) 9667841697 | [kaushikeva0026@gmail.com](mailto:kaushikeva0026@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Medium](#) | [Portfolio](#)

## PROFESSIONAL SUMMARY

Dynamic Data Scientist with a proven track record in driving impactful results through end-to-end data solutions. Expert in Python, R, SQL, and Machine Learning for advanced data analysis and predictive modeling. Skilled in Generative AI, clustering, and database design, with hands-on experience in MLOps and cloud platforms like AWS, Azure, and Databricks. Proficient in Big Data tools (Hadoop, Spark, Hive) and statistical techniques, including regression analysis, Bayesian methods, and time series analysis. Known for exceptional analytical skills and the ability to transform complex data into actionable insights. Eager to contribute to innovative research and teaching in Data Science and Statistics at a leading academic institution.

## 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 & 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.
- Core Member in production of Nes-GPT, implemented 9 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 SDXL Inpainting Model [<https://nesgt.genai.nestle.com/>]
- Conducted 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.
- 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, resulting in reduction in 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, consistently out-performing in rigorous interviews, facilitated my transfer to Analytics Delivery, Data Science at GDN India Capability.

### Founding Pioneer, Daignare

June 2021 - July 2023

- Integrated advanced data science methodologies, including machine learning (ML), deep learning (DL), and statistical analysis, to optimize user experience (UX) and interface design. Utilized convolutional neural networks (CNNs) for image recognition, recurrent neural networks (RNNs) for sequential data processing, and generative adversarial networks (GANs) for synthetic data generation, enhancing user interaction and satisfaction.
- Developed high-precision predictive models and implemented ensemble learning techniques. Conducted comprehensive data analysis using Bayesian inference, regression analysis, clustering, and dimensionality reduction (PCA, t-SNE). Applied advanced statistical methods, including hypothesis testing, time series analysis, and stochastic processes, to derive actionable insights and set new industry benchmarks.
- Engineered and deployed ML and DL frameworks (CNNs, RNNs, GANs) to enhance decision-making in UI/UX design. Improved operational efficiency through advanced optimization algorithms (genetic algorithms, simulated annealing), automated feature engineering, and hyperparameter tuning using grid search and Bayesian optimization. Leveraged cloud platforms (AWS, Azure) for scalable model deployment and real-time analytics, providing a competitive edge in the UI/UX market.

### Advisor, Frontforumfocus Kenya, South Africa

Apr 2021 - Aug 2022

- Implemented 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. Introduced **MLOps** practices to streamline the machine learning workflow, from data preparation and model training to deployment.
- Championed data-driven strategies, increasing decision-making and reducing system bugs by 30%. Enhancing data processing speed by **40%**.

## PROJECTS

- RETAIL-CHURN-AZURE-ML-STUDIO** [[Github](#)]

Developed a predictive model for retail customer churn using Microsoft Azure Machine Learning Studio. Key tasks included data

preprocessing (cleansing, normalization, feature engineering) and implementing a Boosted Decision Tree algorithm. Utilized Python libraries (NumPy, Pandas, azureml-sdk) and PowerBI for detailed visualizations and interactive dashboards, yielding actionable insights for customer retention strategies.

- **MARKOV-CHAIN-MONTE-CARLO-ALGO** [[Github](#)]

Implemented Markov Chain Monte Carlo (MCMC) algorithms for advanced statistical modeling and inference. The project focused on generating and analyzing synthetic data, applying adaptive algorithms, and using Julia for efficient computation. Emphasized data preprocessing, potential function calculations, and diagnostic evaluations to ensure model accuracy and robustness.

- **JENKINS-CODEFLOW** [[Github](#)]

Created a robust automation server using Jenkins for continuous integration and delivery. Developed a Jenkins pipeline with Docker and Kubernetes integration, managing deployments with YAML configurations. Employed Python for scripting and automation, and HTML/CSS for web interface customization, streamlining code deployment and ensuring a seamless workflow.

## SERVICES & 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

**Research Areas:** Advanced Data Science, Adversarial ML, Causal Inference, Reinforcement Learning, GNN, Lifelong Learning.

**Objective:** Developed predictive models for cardiovascular disease risk using logistic regression, SVMs, and neural networks.

#### Contributions:

- Collected and preprocessed large-scale health datasets.
- Enhanced model accuracy through advanced feature extraction.
- Fine-tuned ML algorithms for optimal performance.
- Conducted rigorous statistical analyses and presented findings through detailed visualizations and reports.

**Impact:** Published findings demonstrated ML's potential in early CVD detection, aiding 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**, Perfect A Grade, All India Senior Secondary School Examination (2016)

## TEACHING EXPERIENCE

---

### Data Science Tutor

Superprof (Remote)

July 2022 – Jan 2023

- Instructed an early-stage graduate student from the University of Maryland in advanced Data Science and statistical methods via the Superprof platform.
- Developed and delivered customized lesson plans on machine learning, data mining, statistical analysis, and data visualization.
- Guided the student through complex 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/10.1515/978311398549-010> [[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](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/10.1007/978-99-6698-1\\_4](https://doi.org/10.1007/978-99-6698-1_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](#)]

