Jiyanshu Dhaka

Final Year Undergraduate
Major: Statistics and Data Science
Minors: Computer Science - Machine Learning , Cognitive Sciences

jiyanshudhaka2003@gmail.com
Jiyanshu Dhaka | jiyanshud22@iitk.ac.in
jiyanshud22 | +91-9413186425

ACADEMIC QUALIFICATIONS

Year	Degree/Certificate	Institute	CPI/%
2022 - present	Bachelor's	Indian Institute of Technology Kanpur	**8.1/10**
2021	RBSE (XII)	Disha Delphi Senior Secondary School	**97.8**%
2018	CBSE (X)	Gurukul International School	**94.8**%

SCHOLASTIC ACHIEVEMENTS

- **KVPY(SB)** Scholar 2021 with **All India Rank 28** in common rank list conducted by Indian Institute of Science, Bangalore
- **KVPY(SA)** Scholar 2019 with **All India Rank 832** in common rank list conducted by Indian Institute of Science, Bangalore
- Secured **All India Rank 3846** in **Joint Entrance Exam (Advanced) 2022** among the 0.16 million shortlisted candidates
- Secured **All India Rank 4025** in **Joint Entrance Exam (Main) 2022** among more than 1 million appeared candidates
- Awarded the **Reliance Foundation UG Scholarship**, a merit-cum-means scholarship granted to 5,000 students nationwide
- Achieved **A*** in **Human Centered Computing** and **Computational Cognitive Science** for **Outstanding** performance
- Achieved **A or higher grades** in **14** mandatory or elective courses at **IIT Kanpur** for **Excellent academic performance**

PROGRAMMING EXPERIENCE

- Achieved **Rating of 1600+ (Codeforces Expert)** with **Global Rank 95** in Round 1035 | Profile : **Imvengeance3846**
- Active on Platforms: **LeetCode**: (Jiyanshu_Dhaka) | **Kaggle**: (jiyanshudhaka) | **IB**: (jiyanshu-dhaka) | **GFG**: (jiyanshudifcz)

WORK EXPERIENCE

CHAMPHUNT INC | *Machine Learning Internship* (*Sep'24–Jan'25*)

- Built **Hybrid Post Recommender** (content + collaborative filtering), boosting relevance by **28%** & engagement by **3 min**
- Built **Q-Learning friend recommender** using Q-table updates from user interactions, achieving **78%** user acceptance rate
- Integrated **location match, follower overlap**, & **noise factors** with weight tuning, boosting **personalization** for **5k** users
- Built **Cricket Ball Detection Algorithm** using **Contour Detection + YOLOv8**, **HSV masking & temporal checks**
- Applied **Gaussian blur** & **morphological operation** to reduce noise; Used **Optical Flow** & **Kalman Filter** for smoothing
- Automated extraction of **ball–pitch contact frames; classifying deliveries** as yorker, bouncer, etc. to improve analysis

SALTMINE USA INC | *Workspace Design Automation Intern* | (*Feb'25–Jul'25*)

- Built **Workspace Stacking algorithm** using **greedy allocation**, **proportional distribution**, & **adjacency modeling**
- Generated **85%+ valid** stack plans, with ∼ **90% match** to manual outputs, reducing manual workload by **(∼3–4 hrs/day)**
- Built grid-based zoning engine using **ILP** optimization to satisfy adjacency, periphery, diagonal, and block-pattern constraints
- Implemented **MCMC + combinatorial optimization** based sampling to optimize **objective functions** across constraints
- Developed zoning web app with **NLP-driven interface** translating custom rules into constraints and visualized grid layouts

RESEARCH EXPERIENCE

Passive Image Forgery Detection | *Prof. Nisheeth Srivastava* | *(Dec'23–Feb'24)*

- Reviewed **IEEE** research on passive forgery detection, studying **demosaicing artifacts**, **JPEG traces**, and **CFA** methods
- Implemented **Error Level Analysis** with **HSV contour analysis** to expose tampered regions via pixel-level inconsistencies
- Detected **fake medical scans, X-rays, and morphed reports**, aligning with ongoing **cybersecurity forensics research**
- Classified **507 bonafide** and **210 morph** images as **original or forged**, achieving **80.1%** and **78.7%** accuracy respectively

RL for Odor-Guided Task | *Course Research Project CS786* | *Prof. Nisheeth Srivastava* | (*Sep'24–Nov'24*)

- Modeled an **odor-guided choice task in rats**, where odors signaled forced and free-choice trials, to study task representation
- Developed two **RL models**: **4-state model (generalization across trials)** and **6-state model (odor–action specific)**
- Applied **Hierarchical Bayesian inference with MCMC sampling** on trial-by-trial behavioral data for each animal (rat)
- Conducted **WAIC model comparison**, showing the 6-state model better predicted rat choices with limited generalization

PROJECTS

Cells | *Course Project MTH312* | *Prof. Dootika Vats* | (*Feb'25–Mar'25*)

- Clustered 10,000 cells into 8 distinct types via **spectral clustering**, outperforming **PCA+CCA** by 79.6% in **ARI** metrics
- Achieved the highest **ARI score of 0.87981** in class by integrating multi-omics data using a **Dual-branch Autoencoder**
- Visualized cell latent space with **UMAP**, revealing separation of **cell type** and overlapping modalities across 2 omics layers

EMG Classification | *Course Project MTH209* | *Prof. Subhajit Datta* | (*Feb'24–Apr'24*)

- Developed gesture recognition model using **6 Time-Domain features** per channel, classifying **12 gestures** with **66k** records
- Captured **dual-channel sEMG** signals at **100Hz**, normalized with **Min-Max** scaling, and segmented into **1000ms** windows
- Modeled **SVM (RBF kernel)**, **LR (softmax)**, & **kNN (Euclidean)**; optimized hyperparameters via **gradient descent**
- Applied **PCA** to identify top features and reduce dimensionality; Achieved **94.1%** accuracy with **SVM** & **92.7%** with **kNN**

Generative AI | *GDSC* | (*Dec'23–Jan'24*)

- Learned about **NLP preprocessing**, **Word Embeddings**, **RNNs**, **LSTM**, **Transformers**, and **Fine-Tuning of LLMs**
- Web scraped quotes via **BeautifulSoup** and generated **bag-of-words embeddings** using **NNs** & **co-occurrence matrices**
- Fine-tuned **GPT-2 model** on preprocessed text derived from **word embeddings** to generate high-quality **quotes efficiently**
- Compared performance of **Vanilla GANs (FC/DC)**, **WGANs**, **cGANs**, and **text-to-image GANs** for generating images

Recommendations | *Course Project CGS616* | *Prof. Nisheeth Srivastava* | *(Feb'24–Mar'24)*

- Processed **MovieLens-100k** dataset (943 users×1682 movies), incorporating user demographics, genres, timestamped ratings
- Implemented **item-based**, **user-based**, & **SVD** models; designed **hybrid recommender** integrating similarity+factorization
- Predicted ratings for unwatched movies using **cosine similarity**–based **collaborative filtering** and **SVD**, boosting relevance

Web Surfing Behavior Analysis | *Course Project CGS616* | *Prof. Nisheeth Srivastava* | *(Jan'24–Feb'24)*

- Collected **web log data** to model individual cognitive behavior in browsing, mapping patterns of usage and content interaction
- Applied **Latent Dirichlet Allocation** to extract topic transitions, using **dwell time** to quantify engagement across sessions
- Built **within-day predictive model** estimating likely next-hour topics and visit counts, improving **topic revisit** prediction
- Proposed **stochastic framework** to find **recurrent** and **transitional** websites, predicting future topic switching and revisits

World Demographics | *Course Project MTH208* | *Prof. Dootika Vats* | *(Aug'23–Nov'23)*

- **Scraped demographic indicators** (population, fertility, mortality, GDP) for **115+** countries over **50** years using **R rvest**
- Normalized variables for cross-sectional analysis and built **R Shiny** App, enabling **Data Visualization** in structured manner
- **Data interpretation** using **heatmaps and statistical measures**, identifying **outliers, patterns, trends** and **exceptions**

Intro to Machine Learning | *Stamatics Club, IITK* | *(May'23-Jul'23)*

- Implemented **Linear, Polynomial, and Logistic Regression** models from scratch, achieving $\mathbf{R^2}$ **up to 0.89** on test sets
- Applied **Gradient Descent optimization** to minimize loss functions, reducing training time by **32%** versus naive updates
- Conducted **unsupervised learning** using **KMeans and GMM**, identifying **6 clusters** and uncovering hidden data patterns
- Explored **Neural Networks and CNNs** by generating data, visualizing outcomes, and evaluating **MSE,** $\mathbf{R^2}$ and **accuracy**

Roads | *Course Project MTH312* | *Prof. Dootika Vats* | *(Jan'25–Feb'25)*

- Classified high and variable-resolution road-point datasets as curves or straight segments; achieved **F1 score** of **0.85** & **0.84**
- Derived **geometric features** using circumradius and turning angle from triplets of points to find road **geometry patterns**
- **Optimized** model performance by **grid searching** for **feature thresholds** and evaluating **F1 scores** to enhance **accuracy**

Bayesian Causal Inference in Time Series | *Course Project MTH422* | *Prof. Arnab Hazra* | *(Feb'25–Apr'25)*

- Developed Bayesian multivariate structural **time series model** to estimate causal effects of advertising on sales across stores
- Modeled spatial dependence using **G-Wishart priors** on precision matrix and applied **stationarity constraints** on trends
- Designed a **2-stage estimation**: **EMVS for sparse variable selection** followed by **MCMC sampling** for latent states
- Proposed **Kolmogorov–Smirnov** distance causal estimand, validated via simulations, improving detection of causal impacts

Teaching CNNs Gestalt principles | *Course Project CS786* | *Prof. Nisheeth Srivastava* | *(Sep'24–Oct'24)*

- Implemented **InceptionV4 in Keras**, boosting **CIFAR-10** accuracy by refining feature extraction with custom stem blocks
- Built Python pipeline for symmetry dataset, training **InceptionV3 (TensorFlow)** to achieve **0.55** error rate on **8k** samples
- Trained binary classifier with **InceptionV3 + RMSprop**, applying **early stopping** to reach **0.98 validation accuracy**
- Assessed performance using **ROC/AUC in scikit-learn**, obtaining **AUC = 0.43** on preprocessed grayscale image dataset

Hopfield Networks | *Course Project CS786* | *Prof. Nisheeth Srivastava* | *(Jul'24–Aug'24)*

- Implemented custom **Hopfield Network** model for pattern recognition, achieving 54% accuracy on MNIST data classification
- Optimized network performance via **vectorized weight computation**, reducing computational complexity for large patterns
- Preprocessed MNIST dataset via deskewing, cropping, and thresholding in TensorFlow, boosting **F1** score from **0.06 to 0.52**
- Compared Hopfield classifier to **CNN** and **SVM** using scikit-learn metrics, highlighting **54% vs. 99% and 97%** accuracies

Models of categorization | *Course Project CS786* | *Prof. Nisheeth Srivastava* | *(Oct'24–Nov'24)*

- Implemented **Generalized Context Model** for categorization, using **exponential similarity** to classify **10 test instances**
- Modeled subject bias in **GCM** via **prior probabilities**, capturing **politeness-driven** tendencies in height–weight judgments
- Optimized **feature weights** ($\boldsymbol{\alpha} = [\mathbf{0.65, 0.35}]$), improving sensitivity of predictions and yielding realistic category outputs
- Predicted categories on test sets through **iterative augmentation**, producing stable labels ("2 = average") across scenarios

Learning functions and policies | *Course Project CS786* | *Prof. Nisheeth Srivastava* | *(Aug'24–Sep'24)*

- Implemented **NN for 3-variable Boolean functions** with NumPy, achieving perfect verification across 5 distinct functions
- Developed **Q-learning** agent for Frozen Lake using ϵ**-greedy policy**, optimizing paths with higher rewards over 100k episodes
- Simulated **Rulkov map neuronal dynamics** with custom activation, generating bursting spikes at $\boldsymbol{\alpha}$**=4.5 and** $\boldsymbol{\mu}$**=0.001**
- Evaluated **RL performance via cumulative reward curves**, identifying optimal hyperparameters at $\boldsymbol{\alpha}$**=0.1 and** $\boldsymbol{\gamma}$**=0.9**

TECHNICAL SKILLS

Languages	C | C++ | Python | R | SQL | MATLAB | LaTeX
Libraries	Tidyverse | ggplot2 | tseries | pracma | R Shiny | Rcpp | Rvest | RMarkdown | Quarto | Profvis | NumPy | Pandas | scikit-learn | TensorFlow | PyTorch | Hugging Face | XGBoost | statsmodels | Matplotlib | Seaborn | Plotly
Tools	Jupyter Notebook | Google Colab | RStudio | VS Code | MySQL | MongoDB | Tableau | Docker | Git | GitHub

RELEVANT COURSES

** → A, ** → A**

Data Structure & Algorithm	Introduction to Machine Learning*	AI Techniques in Data Mining*
Data Science Lab (I),(II) & (III)*	Fundamentals of Computing (I) & (II)	Computational Cognitive Science**
Time Series Analysis	Bayesian Models & Data Analysis	Matrix Algebra & Linear Estimation*
Statistical Computing	Linear Regression & ANOVA	Elementary Stochastic Processes(I) & (II)

POSITIONS OF RESPONSIBILITY *(Jul'23–Apr'24)*

Academic Mentor, ICS, IITK	Web Secretary, CWC, IITK |	Executive, Stamatics Club, IITK |