Santanu Biswas		**24N0072**	
Applied Statistics and Informatics		**M.Sc.**	
Indian Institute of Technology Bombay		**Gender: Male**	
		DOB: 09/02/2002	

Examination	University	Institute	Year	CPI / %
Graduation	University Of Calcutta	Ramakrishna Mission Residential College, Narendrapur	2024	8.55
Graduation Specialization: Statistics				
Intermediate	WBCHSE	Ramakrishna Mission Vidyalaya, Narendrapur	2021	93.60%
Matriculation	WBBSE	Ramakrishna Mission Vidyalaya, Narendrapur	2019	95.29%

KEY PROJECTS

Analyzing The Impact of Off-field Factors on Cricket Match Outcomes Using R | *Guided Project*   [2024]

Guide: *Sri Palas Pal, Assistant Professor, Department of Statistics, RKMRC*
- Developed models to analyze the impact of off-field factors on cricket match outcomes and estimate the expected **NRR**
- Performed **Exploratory Data Analysis (EDA)** on the data of over **1030** IPL matches using the **ggplot2** and **dplyr**
- Applied **Backward Elimination** for feature selection and fitted **Multiple Linear** and **Polynomial Regression** models
- Applied **Cross-validation** to ensure robustness, used IPL 2024 results as the test set and **Adjusted R^2** for model validation

Loan Approval Forecaster | *Self Project*   [May'24 - June'24]
- Built classification models to predict whether a loan will be approved or not based on applicants' personal and financial details
- Used a dataset with **11** features and performed **EDA**, removed outliers, encoded variables and performed feature scaling
- Trained models like **Logistic, KNN, Kernel SVM, Decision Tree, Random Forest, Naive Bayes** and **XGBoost**
- **Logistic Regression** achieved the best accuracy of **87.06%** based on **Confusion Matrix, Cross-validation** and **F1-score**

Insurance Charges Prediction | *Self Project*   [July'24 - August'24]
- Performed **EDA** on a dataset comprising **1340** rows and **7** columns, checked linearity assumptions and residual independence
- Multicollinearity was assessed by **VIF** and **Correlation Matrix**, utilized **Gaussian Mixture Models** to get better accuracy
- Fitted models like **Multiple Linear Regression, Polynomial Regression, SVR, Decision Tree** and **Random Forest**
- Performed **Cross-validation** and checked **Adjusted R²**, **Random Forest** gave the best **Adjusted R²** score of **0.879**

Health Assistant: Predictive Diagnosis and Recommendations | *Self Project*   [September'24 - October'24]
- Performed **EDA, Feature Selection, Encoding** with a dataset of 5000 patients consisting of 43 diseases and 132 symptoms
- Fitted models like **Logistic, KNN, Kernel SVM, Decision Tree, Random Forest, Naive Bayes** and **Neural Networks**
- **Kernel SVM** achieved the best accuracy of **89.34%** based on the **Confusion Matrix, Cross-validation** and **F1-score**
- Built a health assistant predicting diseases and providing personalized recommendations with an interactive **Gradio** interface

OTHER PROJECTS

Sentiment Analysis of Financial News | *Self Project*   [October'24 - November'24]
- Processed financial headlines with text cleaning, lemmatization and feature extraction using **TF-IDF** and **Bag-of-Words**
- Trained **FFNN** with different layers, **Naive Bayes, RNN, LSTM** models and optimized hyperparameters with **Grid Search**
- **FFNN** with 2 layers with pre-trained **Word2Vec** embeddings, tuned hyperparameters achieved the best accuracy of **81.03%**
- Developed an interactive **Gradio** interface using the best model for real-time sentiment analysis of financial headlines

Skin Disease Diagnosis System | *Self Project*   [November'24 – December'24]
- Developed deep learning models for accurate **Skin Disease Classification** using image analysis and optimization techniques
- Conducted **EDA** on dataset consisting of **800** images of **9** different diseases, resized and adjusted color channels for consistency
- Performed **Data Augmentation** to fit **CNNs** and some **Transformers**, optimized hyperparameters using **Grid Search**
- **DenseNet** with optimized parameters achieved the best accuracy of **85.48%**, evaluated with different validation metrics

Book Recommendation System | *Self Project*   [November'23 - December'23]
- Processed a dataset of **1,031,175** rows and **19** columns, consisting of review and details of **270170** books by **92107** users
- Performed **EDA**, necessary data manipulations and built **Content-Based Models** and **Collaborative Filtering Models**
- Used **Feature Engineering, Pearson Correlation, k-NN, NLTK, TF-IDF** and **Cosine Similarity** to build the models
- Designed interactive interfaces using **Gradio**, enabling users to receive book recommendations through various algorithms

Hackathon In Cricket Analysis | *Self Project*   [October'22 - November'22]

- Worked on a cricket analysis project for a **Mentored Minds Hackathon**, involving **EDA** and cricketing event predictions
- Processed and cleaned data using **NumPy** and **Pandas**, visualized key trends and patterns with **Tableau** and **Seaborn**
- Analyzed performance metrics and optimized models for better accuracy, contributing to effective insights and predictions

Tesla Stock Price Prediction | *Self Project*   [December'24 - Present]

- Preprocessed and visualized trends of **Tesla stock data (2019-2024)**, calculated returns and analyzed market capitalization
- Conducted **STL decomposition**, stationarity tests (**ADF**) and **ACF/PACF** analysis to identify optimal orders of models
- Built and validated **AR**, **MA**, **ARIMA**, **ARIMAX** and **SARIMAX** models, ensuring reliability through residual diagnostic
- Evaluated models using **RMSE**, **MAE** and **MASE**, finalizing **ARIMAX** for reliable **2-month** Tesla stock price forecasting

POSITIONS OF RESPONSIBILITY

Core Committee Member | *Inferencia 2k24, RKMRC Narendrapur*   [March'24 - April'24]

- Contributed as a key member of the **6-member core committee**, leading the organization of the Departmental Event
- Managed **logistics** and **outreach**, coordinated all event-day activities for smooth execution with faculty collaboration

Design Head | *Inferencia 2k24, RKMRC Narendrapur*   [March'24 - April'24]

- Led the **Design team** of **10** members to create promotional materials and visual content for the departmental academic fest
- Designed **Website interfaces**, event visuals and other event materials, ensuring cohesive branding throughout the event

Sports Coordinator | *Senior Section, RKMV Narendrapur*   [2017-2020]

- Successfully organized all the Intra School cricket and football tournaments in **2017-2018** and the Inter School Cricket Tournament at RKMV Narendrapur in **2018**, also led the volunteer team for Annual Sports in **2017-2020**
- Served as **Coordinator leader** and successfully conducted the Admission Test of RKMV Narendrapur in December 2019

KEY COURSES

- Probability Theory
- Descriptive Statistics
- Statistical Inference
- Sampling Distributions
- Linear Algebra
- Sample Survey
- Time Series
- Data Analysis using R

SCHOLASTIC ACHIEVEMENTS

- Recipient of **DST-INSPIRE Scholarship**, presented by the **Department of Science and Technology, Government of India** for being among the **top 1 percentile** candidates of the 10+2 board exam   [2021 − Present]
- Earned an **Elite Badge** in the NPTEL course on 'Probability and Statistics' provided by **IIT Kharagpur**   [2023]
- Selected as a **Managed Network Expert** at **Chegg India** after successfully passing the exam   [2023]
- Achieved an **International rank** of **681** in the **SOF's International Mathematics Olympiad** in 2018 and **1659** in 2016

TECHNICAL SKILLS

- **Programming Languages** : R | Python | LaTeX | SQL   • **Software** : MS Excel | MS Word | MS PowerPoint | Tableau
- **Interests** : Machine Learning | Deep Learning | Natural Language Processing | Computer Vision | Time Series Analysis
- **Languages** : Bengali | English | Hindi

ONLINE COURSES AND WORKSHOPS

- **Natural Language Processing Course: Deep Learning Bootcamp - Analytics Club, IIT Bombay**
- **Cricket Analytics Workshop - Mad About Sports** in Association with **Rajasthan Royals**
- **Excel Skills for Business: Essential** - Authorized by **Macquarie University** and offered through **Coursera**
- **Fundamentals of Visualization with Tableau** - Authorized by **UC, Davis** and offered through **Coursera**
- **Probability and Statistics** - Authorized by **IIT Kharagpur** and offered through **NPTEL**
- **Machine Learning A-Z: AI, Python and R - Udemy**   • **Data Science with Python - Simplilearn**
- **Introduction to Minitab - Simplilearn**   • **Introduction to Power BI - CareerNinja**

EXTRA - CURRICULAR ACTIVITIES

- Worked in the **Design Team** and **Managed** the event named **INFERENCIA** for three consecutive years   [2021-24]
- **Runner-Up** in the **Inter-Departmental Football Tournament** at **RKMRC Narendrapur** in **(2021-22)**, Received the **Best Goalkeeper Award** in **2022-23** and represented the Department of Statistics for three consecutive years   [2021-24]
- Became **Champion** of the **Inter-Departmental Cricket Tournament** at **RKMRC Narendrapur** in **(2021-22)** and represented the **Department of Statistics** for three consecutive years   [2021-24]
- Served as an **NSS volunteer** and helped organize **Social activities and camps** at **RKMRC Narendrapur**   [2021-24]
- Secured **2nd position** in the **Departmental Chess Competition** of **Mathematics Department** at **IIT Bombay**[2024]