

# Rishabh Sanjay

✉ rishabh.lfs@gmail.com | ☎ +91-8825236258 | in rishabh-sanjay | 🌐 Rishabh261998

## Education

- **Indian Institute of Technology Kanpur** (2017-2021)  
*B.S in Maths and Scientific Computing, Minor in ML and Applications;*  
*Courses:* Data Structure and Algorithms, Introduction to Machine Learning, Linear Algebra, Probability and Statistics  
Probabilistic Modelling and Inference, Statistical Natural Language Processing, Data Mining, Machine Translation, Algorithms II

## Publications

- **ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation** [paper] [code]  
V Malik, **R Sanjay**, SK Nigam, K Ghosh, SK Guha, A Bhattacharya, A Modi  
Association for Computational Linguistics and the International Joint Conference on NLP (ACL-IJCNLP)
- **Semantic Segmentation of Legal Documents via Rhetorical Roles** [paper] [code]  
V Malik\*, **R Sanjay\***, SK Guha, SK Nigam, A Hazarika, A Bhattacharya, A Modi (\*Equal Contribution)  
Natural Legal Language Processing(NLLP) 2022 Workshop EMNLP

## Technical Skills

- **Interests:** Machine Learning, Deep Learning, Natural Language Processing, Competitive Programming
- **Programming Languages:** C, C++, L<sup>A</sup>T<sub>E</sub>X, MATLAB, Python, Java, HTML, JS, SQL
- **Software and Libraries:** Pytorch, Keras, Scikit-learn, Numpy, Pandas, Matplotlib, Bokeh, Pytest, Git

## Research Experience

- **WikiHow QA Dataset Project** (May'22-Present)  
Research Intern : Ohio State University Mentor: Prof. Sun Huan
  - Experimented on downstream tasks to check how **pre-training LM** on the custom dataset before **fine-tuning** on task.
  - Used **RoBERTa-large** for Question Type Classification, Re-Ranking, MRC and Intent Recognition tasks.
  - Used **T5-large** and **bloom** LLMs for **Abstractive Closed Book-ODQA** on the WikiHow FAQ dataset.
- **Court Judgement Prediction, Explanation on Indian Supreme Court Cases** (Mar'20 - June'20)  
Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya
  - Created a corpus of **32,000** legal judgements and tackled the open research problem of Court **Judgement Prediction**.
  - Achieved the best acc. of **78%** (using XLNet) after extensive experiments on 14 SOTA document classification methods.
  - Used different explainability methods such as **IG, Occlusion, Attention etc.** to address explainability.
- **Semantic Segmentation of Legal Documents via Rhetorical Roles** (April'21 - Aug'21)  
Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya
  - Created a new corpus of **100** legal cases, annotated with rhetorical role labels. This is the largest RR corpus.
  - We propose new multi-task learning model **MTL-BiLSTM-CRF(BERT-SC)** which uses label shift as an auxillary task.
  - Our model has an F1 score of **0.71** which is better than various other baselines and sequence classification models.
- **SemEval2021: HaHackathon: Detecting and Rating Humor and Offense** (Sep'20 - Dec'20)  
Course Project: CS771, Mentor: Piyush Rai
  - Developed and trained various machine learning models for Humor classification.
  - Used various ML models from classical models to Transformer based models like **BERT+BiGRU** and achieved F1 of **0.94**.
  - Developed a **BERT+MLP** model to give real number rating to the short text, achieved an RMSE score of **0.446**.

## Work Experience

- **Oracle** (July'21 - September'22) (Hyderabad, IN)  
Applications Engineer, Fusion HCM Development
  - Developed new features for the Oracle Fusion HRMS tool using **Oracle ADF(Java-based Framework)**.
  - Worked on **RESTful** web services for Oracle Journeys.
  - Delivered various **enhancements** and **bug fixes** for both UI and back-end and resolved real-time **customer issues**.
- **Goldman Sachs** (May'20 - June'20) (Bangalore, IN)  
Summer Analyst, IMD Core Engineering Division
  - Created a **Horizontally Scalable** and **Distributed Reconciliation** System in Java for trading data.
  - Used the **divide and conquer methodology** and Akka library in Java to build the distributed system.
  - Queried **Mongo** and **Sybase IQ** databases for the reconciliation process and loaded any inconsistencies.
- **Google Summer of Code at NumFOCUS** (May'21 - Aug'21) (Remote)  
Open-source Developer, ArviZ
  - Implemented plots for Bayesian Visualization like **Dot plots, ECDF plots** etc. with backend as **Matplotlib** and **Bokeh**
  - Used **pytest** library for writing tests and also added **examples** and **documentation** for easier understanding

## Scholastic Achievements

- **KVPY Scholarship Awardee(2017)** by Indian Institute of Science and Govt. of India.
- Secured **All India Rank 2770** in **JEE Advanced 2017** among **220,000** candidates.
- Secured **All India Rank 1633** in **JEE Mains 2017** among the **1.2** million candidates.
- Qualified for **Regional Maths Olympiad(2016)** and appeared for **Indian National Maths Olympiad**.