

Rishabh Sanjay

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

Education

- **Purdue University, West Lafayette** (2023-present)
M.S in Computer Science
- **Indian Institute of Technology Kanpur** (2017-2021)
B.S in Mathematics and Scientific Computing (8.5/10), Minor in ML and Applications(10/10);
Courses: Data Structure and Algorithms, Introduction to Machine Learning, Linear Algebra, Probability and Statistics, Time-Series Analysis, 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
Natural Legal Language Processing(NLLP) 2022 Workshop EMNLP (*Equal Contribution)

Research Experience

- **WikiHow QA Dataset Project** | *Python, Pytorch, HuggingFace, Numpy, Pandas* (June'22-Nov'22)
Mentor: Prof. Sun Huan, Ohio State University
 - Experimented on downstream tasks to check how **pre-training LM** on custom dataset performs before **fine-tuning**.
 - 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.
- **Semantic Segmentation of Legal Documents via Rhetorical Roles** | *Python, Pytorch, Numpy* (April'21 - Aug'21)
Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya, IIT Kanpur
 - 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.
- **Judgement Prediction, Explanation on Supreme Court Cases** | *Python, Tensorflow, Beautiful Soup* (Mar'20 - June'20)
Mentor: Prof. Ashutosh Modi and Prof. Arnab Bhattacharya, IIT Kanpur
 - 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.

Work Experience

- **Oracle** | *Java, HTML, JSF, SQL, Oracle ADF* (July'21 - Sep'22)
Applications Engineer, Fusion HCM Development (Hyderabad, IN)
 - Developed **new features** and delivered various **enhancements** and **bug fixes** for the **Oracle Journeys** product.
 - Worked on the **entire flow**, from designing UI in **Oracle ADF**, implementing and testing the business logic in **Java**.
 - Analysed and resolved real-time **customer issues** every week related to the product.
- **Google Summer of Code at NumFOCUS** | *Python, R, Matplotlib, Bokeh, Pytest, PyMC, Git* (May'21 - Aug'21)
Open-source Developer, ArviZ (Remote)
 - Implemented plots for Bayesian Visualization like **Dot plots, ECDF plots** etc. with backend as **Matplotlib** and **Bokeh**
 - Implemented various algorithms which were incorporated in the implementation of the above plots
 - Used **pytest** library for writing tests and also added **examples** and **documentation** for easier understanding
- **Goldman Sachs** | *Java, Kafka, SQL, MongoDB, Sybase IQ, JUnit* (May'20 - June'20)
Summer Analyst, IMD Core Engineering Division (Bangalore, IN)
 - Created a **Horizontally Scalable** and **Distributed Reconciliation** System in Java for trading data.
 - Created multiple **events** and distributed data so that the data can be processed **concurrently** across the systems.
 - Queried **Mongo** and **Sybase IQ** databases for the reconciliation process and loaded any inconsistencies.

Technical Skills

- **Interests:** Software Development, Machine Learning, Natural Language Processing
- **Programming Languages:** C, C++, \LaTeX , MATLAB, Python, R, Java, HTML, JavaScript, SQL, MongoDB, bash
- **Software and Libraries:** Linux, Pytorch, Keras, Scikit-learn, Numpy, Pandas, Matplotlib, Git, Beautiful Soup, Hugging Face

Scholastic Achievements

- Graduated with a **distinction** in my Bachelors from **IIT Kanpur**.
- **KVPY Scholarship Awardee(2017)** and **INSPIRE scholarship** by Indian Institute of Science and Govt. of India.
- Secured **All India Rank 1633** in **JEE Mains 2017** and **All India Rank 2770** in **JEE Advanced 2017**.
- Qualified in **Regional Maths Olympiad(2016)** and appeared for **Indian National Maths Olympiad**.
- Rating of **1800+** on **Codeforces** and a **5 star** on **Codechef** and solved around **1000+** DSA problems using **C++** online.