# Saisamrit Surbehera

📱 +1 (206) 612 7767
✉ ss6365@columbia.edu
in saisamrit-surbehera
○ saisurbehera

## Academia

**Education**

### Columbia University

| Columbia Engineering ○ Masters of Science in Data Science | ○ Class of 2022 |
| --- | --- |

### University Of Michigan Ann Arbor

| School of Information ○ Bachelors of Science in Data Science | ○ Class of 2021 |
| --- | --- |

Graduated with James Scholar and a GPA of 3.9/4.00.

## Industry

**Walmart**
*2022 and 2023-present*

**ML Research Scientist**, Hoboken, NJ

Algorithmic Re-ranking Team focusing on cold start and boosting
- Implemented an transformer based price prediction model based on item level features including item text, images and engagement
- Developed and deployed multiple cold start pipelines to identify underexposed cold start items and generate embeddings using item information
- Worked on multiple Click Through Rate(CTR) Predictions models fine-tuned for newly added items

**American Family Insurance**
*2021*

**Machine Learning Intern**, Madison, WI
- Created an Information Extraction pipeline (IE) to extract legal definitions and entities from AmFam Policy Form contracts with an 83%. Pipeline turned forms into structured knowledge graphs which is used for answering queries in internal Knowledge App tool.
- Designed and trained multiple co-reference resolution transformer models using semi-structured policy forms.

**Mindgram**
*2021*

**Machine Learning Intern**, Washington, DC

Consulted clients on integrating ML into operations-level decision-making using Palantir Foundry.
- Developed and delivered a Named Entity Resolution (NER) resolution system on SciSpacy to convert unstructured data by fine-tuning a pre-trained BERT model to classify epidemiology-related sentences with an accuracy of 85.5%.
- Implemented a Semantic Role Labelling (SRL) model to understand the latent predicate-argument structure of specific sentences to derive Subject-Object pairs and increased the accuracy to 78.8
- Developed a IE pipeline to extract tables with product sales/revenues of commercialized drugs from SEC filings.

**University of Michigan, Giving**
*2020*

**Data Scientist Intern**, Washington, DC
- Developed an Entity Resolution (ER) system and knowledge graphs for 700,00 living alumni across multiple databases with different level of information for donation campaign targeting
- Integrated Graph Embeddings using DeepWalk / Node2Vec, Tf-idf and word embedding to beat existing internal ER benchmark by 16%.
- Constructed BERT classifier with an API Interface with an accuracy of 92% for identifying underlying topics for all unstructured data (meetings, Text, and email) through domain experience through Wikipedia and relevant books.
- Engineered a full stack search application to find alumni's interests and developed unique knowledge graphs with over 1 million nodes to increase alumni reach.

## Academic Projects

**AAA Insurance**

Implemented ensemble logistic regression, Modified Multi-variate Gaussian (MVG), Modified Randomized Under sampling (MRU), and Random Forest to select the best model to predict the number of breakdowns for a model in a zip-code with 85% accuracy

**University of Stellenbosch**

Lead a team of four members for over nine months to build a re-modeled business conflict barometer, an automated and analysis tool of textual data and variables within date, location, etc. to determine when and how conflict levels are rising or falling based on a given situation. Pipeline included several scrapers and multiple BERT models to predict conflict scores

## Skills

**Technical**

General Programming (Java, Python)  ○  Data (R, SQL, PySpark)  ○  Deep Learning (PyTorch)