

# Ankit Sharma

## Senior ML Engineer, Shopee

Singapore

+65-88254010

ankit19910508@gmail.com

www.linkedin.com/in/ankitsharma387

Senior Machine Learning Engineer having **6+ years** of industry experience in ML, NLP, deep learning, feature engineering, big data analysis and information retrieval along with data structure and algorithms. Specialized in using AI to improve search and recommendation for e-commerce products. Developed and shipped full-stack ml-based solutions which resulted in significant business impact and user experience. Known for goal-oriented problem-solving, analytical thinking and data-driven development.

## Work Experience

### Shopee, Machine Learning Engineer

Nov 2019 - Current

#### Query Category Prediction

- Classification of user query into the existing Shopee category tree to improve search relevance.
- Developed a end-to-end **ML pipeline** that includes **feature engineering**, training **GBDT model** & **A/B testing**.
- Designed and implemented an **active learning** strategy to identify incorrectly labeled data and retrain the model with modified weights to avoid the need for remarking.
- Did offline inference and stored query-category pairs in cache for **latency optimisation**. Enhanced coverage for long tail queries by leveraging the existing drop word service.
- Impact:** Average **decrease in bad case rate by 4.5% & 2.5% OPU increase** across Southeast Asia.

#### Item Tagging

- Tagged items with product type for richer indexing and search result improvement, trained **CatBoost model**.
- Designed features including **cross features** of item & product type using **BERT embeddings**, this boosted accuracy. The output product type tags are also used by many downstream projects
- Impact:** Average **decrease in bad case rate by 3.8% & 1.8% OPU increase** across Southeast Asia.

#### Item Accessory Classification

- Distinguish accessories from main item by training **GBDT model** using **tf-idf**, price cluster & category as features.
- Impact:** Average **decrease in bad case rate by 1.2%** across many regions in southeast asia.

#### Miscategorized Item Detection

- FAISS's approximate nearest neighbor** and **heuristics** based on price, user behavior, and **embeddings** were used to identify items with incorrect categorisation by sellers.
- Impact:** **Bad case rate decreased by 1.1%** in Singapore market.

### Concerto AI, Machine Learning Scientist

Jan 2019 - Oct 2019

#### Patient's health data prediction

- Predicted survival timelines for cancer patients using **XG boost** trained on their historical data.
- Predicted Cancer & Smoking status of patients with **85% accuracy** by **finetuning ULMfit model (transfer learning)** on their prescriptions.

### Dataweave, Data Scientist

Jan 2017 - Jan 2019

#### Catalog comparison across different e-commerce company data

- Classification of different e-commerce products into product types generated from taxonomy using **SVM**.
- Match similar product across e-commerce websites using **embedding based ranker** and heuristic.
- Used **GPC and Google taxonomy** to create global taxonomy for Dataweave.
- Successfully delivered this product output to 20 client companies and got **appreciation from the CEO**.

### Cisco, Network Consulting Engineer

Aug 2016 - Dec 2016

- CCNA Certification in Routing; Switching and Data Center

## Education

Degree	Institution	%/CGPA	Year
MTech (Computer Science)	Indian Institute of Technology, Madras, India	8.34	Jul 2014 - Jul 2016
BTech (Computer Science)	M.I.T Pune	6.3	Jul 2011 - May 2014

## Skills

- Technology:** Python, Golang, Scala, Scikit-learn, Numpy, Pandas, Jupyter, Gensim, Pytorch, Spacy, Fasttext, NLTK, Zeppelin, MYSQL, NOSQL, HDFS, Map-reduce, Elasticsearch, Kafka, Linux, kibana, EFK, Nginx, CherryPy, Flask
- Major Course work:** NLP, Machine Learning, Artificial Intelligence, Data mining, Advance Programming Lab

## Positions of Responsibility and Achievements

---

- Publication: "Case Representation & Retrieval Techniques for Neuroanatomical Connectivity Extraction from PubMed"
- Secured All India Rank 105 among 1.55 lakh applicants in GATE 2014.
- Selected in Onsite round of ACM-ICPC(2013) programming competition.