Nissincho 23-01, #911
Kawasaki City, Tokyo, Japan

Avinash Mishra

+81-70-2188-8241
avinash.mishra.2388@gmail.com

EXPERIENCE

Data Scientist	*Catalina Marketing, Japan*	*June. 2018 ~*

NPA
- New Product Accelerator is a sales forecasting tool for FMCG products. It is powered by a powerful model ensembled with machine learning and deep learning. It also has a powerful forecasting dashboard.

ID mapping
- Id mapping uses Locality sensitive hashing algorithm to match real customers using their geo-location data.

Likely buyer
- Prediction model to find customers, it targets likely buyer customers to print a coupon for them.

Product2Vec
- It is a product similarity model. It Uses word2vec to find a similar product at basket level.

Learning to reRank
- This is a machine learning model which uses CTR to rerank offers in our c-wallet app.

Customer Segmentation
- Various customer segmentation based on shopping behaviour like Working Mama, Early Adapter, Digital literacy, Like travelling etc.

Data Scientist/Team Lead	*Paradigmshift.io, Japan*	*Sept. 2016 ~May 2018*

Valuechain [link] [see project]
- Valuechain is a machine learning tool which provides deep insights into OTAs reviews.
- I have designed the architecture and we perform various NLP techniques like Noise removal, Lexicon Normalization, Object standardization, Entity Parsing, Statistical featuring, text classification and matching etc.
- Tools and technologies used: NLTK, MeCab, ML-Classifier, Python, Flask.
- Managing AWS instances of Valuechain server.

Price RecSys [link]
- This is a price recommendation engine for OTA price master. We are using Hybrid filtering with Spark ALS to recommend the price.

Smart web crawler [link]
- Designed distributed web crawling architecture. It has a centralized controller, distributed spiders and distributed storage in Cassandra cluster.
- Tools and technologies used: Apache Kafka, Scrapy web crawling framework, Apache zookeeper, Django, Python, bash.

Software Engineer(Big Data Department)	*Rakuten, Japan*	*Oct. 2014 - Sept. 2016*

GSP [see project]
- GSP is Global search platform which is being used for almost all of the Rakuten services.
- Created an automated testing framework (Ngauto) for distributed search services involving systems like Solr, Zookeeper, Cassandra and Hadoop.
- Created search peripheral components like dictionary compiler and word extractor to support search sub-functions like "did you mean", "related-words", "spell-check" and "auto-completion".
- DevOps: In-charge of CI/ CD. Used docker, chef, Jenkins and some shell scripts for OS-provisioning and delivery pipeline.

Survey Panda

- This was our training project we as a team designed Survey Panda for PC support help desk feedback. The project was evaluated as the best training project of October 2016 batch and the product was chosen to be used at the PC support help desk.
- I was part of backend team working on Spring boot (J2EE). My main task was to design a survey database and validate survey form and post feedback form.

Dynamic Search UI [see project]
- Initial team member responsible for improving dynamic search UI of Rakuten ichiba.

Systems Engineer	Infosys Ltd., India	Mar. 2011 - Aug. 2012

Molina Healthcare management
- Molina was a client of Infosys and they wanted to upgrade their healthcare product. It was a pretty big project and we were a team of 20+ engineers. I was responsible to manage users' health-related information in the database. I designed the DB in SQL server and performing stored procedures, triggers and functions to optimize it.
- Technologies: Java, Python, SQL server.

Employee SWAP portal
- India with a big demographic region sometimes people get posting at a place where they don't want to live. So my team and I have designed one web-based swap portal for employees who are interested to swap their posting locations. It made the transfer process easy.
- Technologies: Java, JSP, Servlet.

Teaching assistant	VJTI Mumbai, India	Sept. 2013 - June 2014

During my Master of Technology course, I was working as a Teaching assistant in VJTI for various courses mentioned below.
Courses: Data structures and algorithms, Software engineering, the Mathematical foundation for computer science.

EDUCATION

Master of Technology	Software Engineering, Department of CS (VJTI), Mumbai, India	Sept. 2012 - June 2014
Bachelor of Engineering	Computer Science & Engineering (RGPV), Bhopal, India	July 2006 - June 2010

FUN PROJECTS

For my Fun Projects, I create GitHub repositories to convey some vertical ideas.
- *NLP-with-NLTK* [see project]
- *Cassandra snapshot Backup and restore* [see project]
- *Aspect-term-extractor* [see project]

ADDITIONAL EXPERIENCES/ AWARDS/ HOBBIES

- Rakuten Project Award for GSP.
- Event organizer in Rakuten Technology conference 2014.
- Placement coordinator of software engineering branch during 2013-2014
- **Hobbies:** Solving puzzles, Reading about International affairs, Playing Badminton.

LANGUAGE AND TECHNOLOGIES

- Language: Java, Python, SQL, bash, Scala, JavaScript, PHP
- Server: Apache Tomcat, Jetty, Nginx
- Big data technologies & tools: HDFS (HDP), Hive, MapReduce, Spark, Kafka, Zookeeper, Ambari, AirFlow
- Search technologies: Apache Lucene, Solr
- Database: Mysql, Cassandra (Apache, DSE), MongoDB, MariaDB, SQL Server, Redis, Netezza
- DevOps/CI : Jenkins, chef, docker, Vagrant, Apache Maven
- Tools: GitHub, Jira, Confluence, Bitbucket, Gitlab, target process
- Web Frameworks: Django, Flask, Laravel, VueJs, Node.js
- OS: Ubuntu, CentOs, Amazon AMI, macOs
- Data Science: NLP, NLTK, Scikit-Learn, Keras, Tensorflow, Spark MlLib, Mantra
- Data Visualization tools: Tableau, DOMO, incubator-superset, matplotlib, Dash by plotly
- Cloud Platform: AWS, Digitalocean, Amazon Lightsail