RITESH RATTI

Phone:(+49)1744363742 WhatsApp:(+49)1744363742 | E-Mail: ritesh.ratti@gmail.com

LinkedIn Profile: https://www.linkedin.com/in/ritesh-ratti-a5311615/
Google Scholar Profile: https://scholar.google.com/citations?user=LgOa2lsAAAAJ&hl=en

Applied Data Science & ML / Deep Learning / Text Analytics / Big Data

Professional Summary

- 13+ years of professional experience in applied **Machine Learning & Data Science**.
- 2+ teams led successfully as **Technical Lead**.
- Architected, Engineered and Delivered intelligent software solutions using machine learning techniques.
- Experience on leading machine learning based software projects.
- Experience in statistical machine learning and predictive modeling to drive insights and recommendations.
- Strong capabilities in Classification and Clustering techniques.
- Strong expertise in **Data Mining and Text Analytics**.
- Experience in developing high performance and scalable data pipelines for distributed processing.
- Published research papers in top conferences in the area of Machine Learning and Network Security.
- Quick Learner of new platforms, technologies, and rapid prototyping.

Professional Experience

Delivery Hero, Berlin, Germany | Senior Data Scientist | Food Data Science Team | March 2022 - Present

Roles and Responsibilities

- Leading team of 3 data scientists and managing projects for the food data science team.
- Technical leader for food science projects and working closely with regional teams to deliver the models and ML solutions.
- Responsible for setting quarterly OKR's along with project manager and setting up the roadmap of data science projects and technical delivery of projects.
- Participate in deep technical architecture discussions about design and code review PR's.
- Actively involved in product development and mentoring peers during the development lifecycle.
- Leading the development of Vision related DS projects like Dish Image Classification, Placeholder detection etc.
- Designed and Developed Food item category and sub category prediction through multilingual Distil Bert model for 9 EU and 5 APAC countries using Transformers and achieved weighted f1-score of 80%. Also developed the mechanism for ranking vendors through predicted dish attributes from the model.
- Developed POC on using Graph analytics based Community Detection for customers based on predicted Dish Attributes.
- Developed multimodal mechanism for dish classification using text and image together with early fusion technique using TensorFlow. Improved metrics by 4-5 % using Bert and Mobilenet based models.
- Food Image classification models using Mobilenetv2 and EfficientNet techniques covering 500 classes with accuracy of 65% and Placeholder image detection using Siamese networks with Contrastive and Triplet loss functions.
- Utilizing Nlpaug specific data augmentation technique to include different dish names scenarios.

Platform/Languages/Tools: Python, Transformers, Scikit-learn, Tensorflow, Docker, Kubeflow, GCP (google cloud platform).

Grab, Singapore | Senior Data Scientist | Grab Food Data Science Team | Nov 2018 - Dec 2021

Roles and Responsibilities

- Actively involved in product development and mentoring peers during development lifecycle and building the roadmap for software tools along with manager and product owners.
- Architected and developed solutions using text analytics and computer vision algorithms.
- Leading the development of computer vision related data science projects like Image correction, Mart Item Tagging, etc. and leading the development of Eater side recommendation projects using data science algorithms over big data.
- Designed and Developed Food Mart Item category and sub category prediction through Distil Bert model for 6 different countries using transformers based multi-label multi-class classification model with average accuracy of 70%..
- Developed Food-Nonfood detector system using transfer learning and Mobile Net v2 modeling using TensorFlow with accuracy of 75%.
- Developed Watermark detector model using ResNet 50 modeling supporting multiple watermarks and developed object detection model for Watermark using YOLO-v3 model that improved the accuracy by 2-4 %.
- Developed up-selling items recommendation system using FP Growth Algorithm using Spark MLLib for millions of customers.

- Developed Trending Recommendation Learning to Rank Model to identify the geo hash specific ranks for distance / popularity score / fulfillment rate etc. using logistic regression based model that improved NDCG score by 1-2 %.
- Developed statistical insights on promotions and menu items performance on data lake using Spark that is currently used by more than 300K merchants.
- Estimation of Food preparation time based on Button Press through apps at various stages using statistical methods.
- Implementation of batch job processing using airflow scheduling functionality for all the machine learning projects.

Platform/Languages/Tools: Python, Spark, Airflow, Scikit-learn, Spark MLLib, Keras, Tensorflow, Transformers

Pitney Bowes Software, Noida, India | Senior Advisory Software Engineer | Apr 2016 - Oct 2018

Roles and Responsibilities

- Leading the architecture development of the project along with a team of engineers and project managers and interacting with stakeholders and defining their requirements in terms of actionable development stories. Also mentoring peer engineers to make sure development of the product is on the right track.
- Designed & Developed Smart Data Quality solution using Active Learning Technique based on clustering and classification and conducted research on various techniques of clustering methods for grouping similar records for deduplication.
- Implemented Offline clustering based on LSH (Locality Sensitive Hashing) and BIRCH Clustering techniques and Developed variation extraction scheme using generated clusters to implement active learning methodology.
- Development of automatic Rule Extraction scheme for match key and match rule generation through Decision Trees and Development of Metadata utility in address parsing process to get the information at every step of execution.
- Leading the development of Address parser framework and designing various modules for processing pipeline. Also developed a Multi-Layer Perceptron based address parser for UK and Germany with Word2Vec vectorization over Global Address data for Classification purpose that achieved an accuracy of 75%.
- Developed POC on LSTM based model for address parser, modeling as a sequence prediction using learned embedding that improved the overall score 2-4 %.
- Performance Optimization of machine learning code and optimum hyperparameter selection for machine learning models.
- **Published research paper in conference ICSC 2018 at Los Angeles, U.S**
- Designed & Developed Entity Extractor module for extracting entities like Person, Location, Organization, Address, Phone Number etc. using Conditional Random Field (CRF) model and building ML models over CoNLL-2003 and MASC data set & comparing the results with existing solutions.
- Developed Text Categorization functionality for custom data using SVM & MaxEnt model.
- Designed & Developed Entity Extractor for Bank wire text used for clients in the Banking domain.
- **Published & Presented research paper on Bank Wire Entity Extraction in IJCAI SML 2017 at Melbourne, Australia**

Platform/Languages/Tools: Java, Python, Scikit-learn, Gensim, Keras, Cleartk, Mallet, Spring, Maven.

Airtel Wynk, Gurgaon, India | Senior Software Engineer | Wynk Backend Team | Feb 2015 - Apr 2016

Roles and Responsibilities

- Involved in the initial feature roll out on songs recommendation system and developed an asynchronous event processing framework using storm and kafka that improved the TPS to 5x times.
- Developed Songs Recommendation system using Storm-trident based aggregation functionality.
- Designed and developed Dynamic My Favorites section based on user activity. Different user activities are scored differently and analyzed over Hadoop infrastructure and prioritize the user content based on generated Hive reports.
- Integrated google now functionality to create context specific cards. Cards are created based on user activity and New album release cards are created for all users.
- Design & Development of event processing framework using Storm and Kafka and Implementation of Spouts and Bolts for stats processing through Storm.
- Designed scalable and fault tolerant production architecture for event processing and deployment over AWS.

Platform/Languages/Tools: Java, Apache Storm, Apache Kafka, Zookeeper, Mongo DB, Maven, Linux.

Samsung Research, Bangalore, India | Technical Lead | Cloud Computing Lab | Dec 2012 - Jan 2015

Roles and Responsibilities

- Actively involved in advanced feature development using Text mining algorithms.
- Design & Development of common framework for Feature Extractor using text analytics algorithms over Hadoop based infrastructure.
- Implemented various text analytics features like Tf-Idf, Topic modeling, entity extraction over Hadoop infrastructure indexed on Elastic Search for Information retrieval.
- Design & Implementation of Topic extraction from Unstructured data files using LDA Model and Q/A Scoring technique is devised using extracted topic probabilities.
- Designed the Map Reduce based workflow for Feature Extractor and implemented the workflow in Oozie.
- **Automated complete flow for Offline processing using Oozie at Samsung Headquarters, Korea**

- Designed and Developed CaaS (Configuration as a service) infrastructure components and implemented camera use cases for providing real-time configuration suggestions.
- Implementation of POS Tagging functionality for identifying actions from voice based commands.
- Implementation of Clustering based solution for intelligent configuration suggestion using K-Means Clustering.
- Design & Implementation of Content Management for Social Service Framework by providing various transcoding schemes and metadata extraction from media files. Developed api for content upload and download .
- Developed caching scheme using Memcached that improves the access performance of objects by 20%.
- **Published Patent on Cross Region Caching based approach.**

Platform/Languages/Tools: Java, Spring, Weka, Memcached, ffmpeg, HBase, Hadoop, Oozie, Mallet, Zookeeper.

Oracle India, Bangalore, India | Member Technical Staff | Database Security Team | Jul 2010 - Nov 2012

Roles and Responsibilities

- Development of multiple log collectors and fixing critical issues in different product releases.
- CLI development for easy administration of product functionality.
- Design and development of OSAUD, DBAUD Collectors and CSDK according to new push based Event Model and Implemented OSAUD, XML and SYSLOG Collectors for AV11g.
- Implementation of logging scheme and Information Lifecycle Management for Collector Metrics.
- Designed and developed command line utility AVCLI to provide user friendly console for interaction and processing and Implemented Firewall Management and Administration through AVCLI.
- Modification of Make files and Plug-in formation and fixing critical issues in AVCLI.

Platform/Languages/Tools: C, Java, OCI (Oracle Call Interface), XML, Linux x64.

Education

2015 - 2023 Phd. (Part time) Computer Science and Engineering, Indian Institute of Technology, Guwahati, India.
      Dissertation Submitted: - *Application specific Network Intrusion Detection System (An unsupervised perspective)*

2008 - 2010 M.Tech. Computer Science and Engineering, Indian Institute of Technology, Guwahati, India.
      Dissertation Topic: - *Active Detection mechanisms for attacks in Autonomous Systems*
      Bandwidth Measurement. **CPI: 9.1/10 [Published & Presented research paper in ISA 2010 at Miyazaki, Japan**

Technical Skills

Programming Languages	Python, Java
Machine Learning Technologies	Tensorflow, Transformers, Scikit-learn, OpenCV, Keras, Mallet, Gensim, Nltk
Big Data Technologies	Spark, Hadoop, Kafka, Storm, Zookeeper, Docker, Kubernetes
Data Storage Technologies	DynamoDB, MongoDB, Redis, HBase, Memcache, Oracle PL/SQL, Presto
Monitoring /Scheduling Tools	Airflow, Kibana, Datadog, Wireshark, Oozie

   I hereby declare that all the above information is true to the best of my knowledge and belief.