Aakash Goel

[Analytics Diksha YouTube Channel](https://www.youtube.com/channel/UC4yh4xPxRP0-bLG_ldnLCHA) [Medium](https://medium.com/@aakashgoel12)

LinkedIn <https://github.com/aakashgoel12>

**Experience (4.2 Years)**

**Senior Data Scientist** @ **Fractal Analytics Pvt. Ltd.** **May 2017 – Present**

Project: Customer care's E-mail/call Analysis for a retail client

* Developed complete analytics framework that helps in identifying the reason of contact by customer
* Tools/Techniques: Ensemble of Multi-label classification (Logistic Regression, LinearSVM, Decision Tree),

Singular Value Decomposition, Word Embeddings, Google cloud, Speech-to-text Project: Predicting Coupon Redemption

* Help retailer’s marketing team to effectively design coupon construct and precisely measure consumer’s propensity towards coupon usage
* Technique: Tree based boosting algorithm model (Catboost) based on the customer, campaign and item level features**; 0.88 AUC score**

Project: Text Similarity

* Given two sentences, find if they are Semantically identical or duplicate
* Tools/Techniques: Developed Siamese deep neural network (CNN, LSTM). Involved Text preprocessing,

similarity distances of various embeddings, syntactical features; Python, Keras Project: Catastrophe events prediction for Insurance client

* Customers impacted by catastrophe (CAT) events claims millions of dollars from insurers. To mitigate impact of claims, need to precisely identify claims associated with CAT events
* Technique: Dealt with highly imbalanced dataset (structured and unstructured data); classification model (Catboost, Random Forest)**; 0.92 AUC score**.

Project (R&D): Unsupervised Learning in Text

* Entity Extraction: Given seed Entities and Corpus, expand entities of specific type in unsupervised manner, used Bootstrap Pattern Learning; Supervised – CRF
* Aspect/Opinion Mining: Given only review containing multiple opinions for product/service, get sentiment for each opinion; used double propagation method, Topic Modelling
* Event Extraction: Building event database from News, used Context Extraction
* Information Retrieval System: Used BM25 to recommend Vitamins based on user’s query(disease/symptoms)
* Entity Resolution: Mapping of entities present across different databases to find prospects/Opportunities

**Research Software Engineer** @ **Monster India Pvt. Ltd.** **October 2016 – May 2017**

* Resume Parser: Designed scoring system to extract entities from education section using lexicon, contextual words and rule based
* Used Locality Sensitive Hashing to find similar organization names

**Research Staff @ IIT - Delhi** **June – October 2016**

Lab: SCFBio (Supercomputing for Bioinformatics) Laboratory

* Database Management, MySQL
* Algorithm study – String Matching.



**Technologies & Tools**

**Techniques:** Supervised Techniques (Regression, Classification), Unsupervised techniques (Clustering, DimensionalityReduction, SVD, PCA), Ensemble modelling - bagging and boosting techniques like Random Forests, XGboost, Catboost. Deep Learning (CNNs, RNNs, LSTMs), Text Modelling (Dialog Systems - chatbot, Topic Modelling, Entity Extraction, Embeddings, Recommendation Systems)

**Tools (Recently worked in)**: Keras, Python (Sklearn, pandas, numpy, NLTK, Stanford NLP Tools, Spacy, Gensim,Pattern), SQL

**Publication**

* Dr (Prof.) Saroj Kaushik, Sunita Tiwari, Chhavi Agarwal, **Aakash Goel** **–** “Ubiquitous Crowdsourcing Model for Location Recommender Systems” in International Conference on Computer Technology and Development, Singapore, 2015. Publisher: Journal of Computers [(JCP, ISSN: 1796-203X, doi: 10.17706/jcp.11.6.463-471](http://www.jcomputers.us/index.php?m=content&c=index&a=show&catid=181&id=2662)).

Email: aakashgoel12@gmail.com Ph: +91-9953456627 (M)

**Academic Details**

|  |  |  |  |
| --- | --- | --- | --- |
| B.Tech. (Computer Science & Engineering) | JMIT, Radaur (Kurukshetra University) | 75% | 2016 |
| All India Senior Secondary Class XII | D.A.V Centenary Public School, CBSE Board | 84.8% | 2012 |
| All India Secondary Examination Class X | D.A.V Centenary Public School, CBSE Board | 8.2 CGPA | 2010 |
|  | |  |  |
| **Research Internships during Engineering** | |  |  |
| • “Tourist Spot Recommender System using Crowdsourcing Concept”, **IIT** **–** **Delhi** | | **June – July 2015** | |

Objective: Generate Quality recommendation of places close to tourist’s current location using crowd sourcing approach. Contextual information is collected from crowd in terms of fuzzy linguistic variables to generate a popularity score of each place nearby tourist’s current location.

* “Software Authentication Utility for Defense Applications”, **Defense Research and Development**

**Organization (DRDO, Delhi)** **July – August 2014**

Objective: Developed software to protect DRDO system from unauthorized access. It fetches the MAC address of the system stored by the system itself during its first execution, Encryption of MAC address is done & this encrypted value is stored in window’s registry. On each subsequent execution the system matches the registry values with the present value entered by user to perform the successful login.

**Relevant Projects/ Competitions**

* **Flight Delay Prediction:** Predict whether flight is going to delay or not. If flight delays, predict amount of time bywhich it delays [**(**https://bit.ly/2wpuDc9](https://bit.ly/2wpuDc9))
* **HackerEarth Machine Learning Challenge #6 (Predict damage to building):** Worked on predicting degree ofdamage that is done to building post an earthquake. Implemented Catboost classifier for multiclass problem and ranked 28th out of 6000+ participants with weighted F1 score 0.79137 [(https://bit.ly/2u4UZze)](https://bit.ly/2u4UZze)
* **Linguipedia codeFest Natural Language Processing (Sentiment Analysis):** Linear SVM for Sentiment analysis ontwitter data; Shingles and word segmentation for hashtags. Ranked 17th with 0.8974 F1 score [(https://bit.ly/2vK364A](https://bit.ly/2vK364A))
* **Kaggle Toxic Comment Classification Challenge:** Implemented Logistic Regression with words and char n-grams.Objective was to build a multi-headed model capable of detecting distinct types of toxicity like threats, obscenity, insults, hate. Achieved ROC AUC 0.975 score on Kaggle.

**Areas of Interest**

• Natural Language • Machine • Deep • Open Information • Programming

Processing Learning Learning extraction

**Awards & Achievements**

* Delivered Talk in Kaggle Meetup on Semantic Text Similarity [(https://bit.ly/2SY6AJ4](https://bit.ly/2SY6AJ4))
* **Courses**
  + Udacity Machine Learning Engineer Nanodegree [(https://graduation.udacity.com/confirm/GKC99A7N](https://graduation.udacity.com/confirm/GKC99A7N))
  + Improving Deep Neural Network (<https://bit.ly/3g75DIx>)
  + NLP in TensorFlow (<https://bit.ly/330HG22>)
  + NLP with classification and Vector Spaces (<https://bit.ly/3jICeGQ>)
  + Introduction to Practical Deep Learning (<https://bit.ly/3f5czVt>)
* Awardee of Prestigious Indian Institute of Technology Delhi Summer Research Internship Award (2015)
* Institution’s Team Leader for “National Entrepreneurship Challenge” organized by **e-cell, IIT Bombay**, 2013, Team ID: ECELL0254
* “Top 10% Merit Certificate” in National Standard Examination in Physics (2011)
* Secured 8th Position in City in 12th National Science Olympiad. (November 2009)

|  |  |  |  |
| --- | --- | --- | --- |
| **Academic References:** | |  |  |
| \*Prof. Saroj Kaushik | | \*Dr. Gaurav Sharma | |
| HOD, CSE Dept, IIT Delhi | | HOD, CSE Dept, JMIT | |
| Email-id: [saroj@cse.iitd.ernet.in](mailto:saroj@cse.iitd.ernet.in) | | Email-id: [gauravsharma@jmit.ac.in](mailto:gauravsharma@jmit.ac.in) | |
|  |  |  |  |

Email: aakashgoel12@gmail.com Ph: +91-9953456627 (M)