Sumit Sidana

Contact Information

Senior Data Scientist at Takeaway.com
Am Karlsbad 16, 10785 Berlin– Germany
sidana.sumit@gmail.com

Website: https://sumitsidana.github.io

sidana.sumit@gmail.com
(+49)15207500582

Internships and organizational experience

- **"Large-scale Recommender Systems and Ranking at JustEat Takeaway.com"**
  Senior Data Scientist (2019 - Present)
  - Working on building prototypes for personalized restaurant and dish recommendations in order to rank items according to customer's preference
    - Built and productionalised implicit matrix factorisation for personalised restaurant recommendations using SageMaker.
      - Used to replace popularity based baseline in personalized restaurant emails
    - Built a POC for personalised restaurant recommendations using Factorisation Machines
      - Attempt to replace strong personalized restaurant recommendations baseline on multiple channels
    - Participated and presented in recommendation meetups inside the company, which dealt with problems related to various biases and unbiased learning to rank.
    - Literature review on counter factual offline evaluation using multi-armed bandits and inverse-propensity metric.
  - Building a relevance search prototype using Learning to Rank (LTR) principles
    - Query Expansion using item2vec models
    - Rank boosting using position-bias corrected click-through-rate
    - Well versed in learning to rank loss functions
    - Involved in A/B testing newly introduced features
    - Tools Used: Elastic Search
- **"Building large scale recommender sytems leveraging implicit feedback"**
  Ph.D. student at Laboratoire d"Informatique de Grenoble (LIG), under Prof. Massih-Reza Amini (2015 - 2018)
  - Worked under the supervision of Prof. Massih-Reza Amini and Post. Doc. Charlotte Laclau in order to apply deep learning methods to improve top-k recommendations on implicit feedback
    - Tools used: Tensorflow
    - More details here: https://arxiv.org/abs/1705.00105
  - Contributed a dataset consisting of click logs of users of Kelkoo for recommender system community
  - Set benchmark results for major recommender system baselines meant for implicit feedback on the above mentioned dataset
    - More details on: https://dl.acm.org/citation.cfm?id=3080713
  - Working in collaboration with engineers at Kelkoo and Purch on the project Calypso where the objective is to predict the probability that user clicks on a given offer using Field-Aware Factorization Machines (FFM) and feature engineering.
    - Data pre-processing in SPARK
    - FFM implementation taken from https://www.csie.ntu.edu.tw/~cjlin/libffm/
  - Attended RecSys 2016 held in Boston and RecSys Summer School 2017 held in Bolzano
  - Got papers accepted at A* conferences (SIGIR) and core-A journal (TKDE). More details here: dblp, Google Scholar

- **"Mining health and nutrition information from Twitter over time"**
  Research Intern at Laboratoire d"Informatique de Grenoble (LIG), under Dr. Sihem Amer-Yahia (May 2015 - October 2015)
  - Worked in collaboration with Profs. Marianne Clausel and Massih Reza Amini to extend LDA with hiddden temporal variables in order to discover seasonal diseases in Twitter. More details here: `https://ieeexplore.ieee.org/document/8263414/`
  - Ported the data pipeline developed in the first internship onto ShareInsights, a robust data manipulation infrastructure
  - Found general-purpose concepts like retail products and diseases inside French tweets and statistically relating them

- **"Mining health and nutrition information from Twitter"**
  Research Intern at Laboratoire d"Informatique de Grenoble (LIG), under Dr. Sihem Amer-Yahia (September 2014 - December 2014)
  - Worked in collaboration with computer scientists and geographers in the context of the CNRS MASTODONS CrowdHealth project where I developed:
    * A database indexing module in Postgres to optimize tweets extraction in real time
    * A tweet annotation module based on crowdsourcing:
      `https://crowd4u.org/en/projects#p5`
      `http://nutritionunleashedforus.com/sidana/user/nutritionCrowd.html`
    * An SVM classifier module based on a 10-fold cross validation
    * A Gibbs sampler module for inferring health information via sophisticated bayesian modelling
    * Developed a geo-based visualization interface
  - Project details available at http://slide-apps.imag.fr/crowdhealth/

- **"Feature engineering and implementation of machine learning algorithms at Xurmo Technolgies Bangalore "**
  Research Engineer at machine learning based product-based start-up Xurmo Technologies,Bangalore (June 2013- November 2013)
  - Studied and wrote jobs for filtered and wrapper based feature selection methods
  - Algorithms implemented - Information gain, Chi-Square, Principal Component analysis and forward feature selection algorithm (in java)
  - Ran the algorithms on many UCI-repository datasets.
  - Results were at par with popular data mining tool - Weka.

- **"System Engineer at Estel Technologies "**
  System Engineer at Global Support Team at ESTEL TECHNOLOGIES PVT LTD Gurgaon, India (February 2009 - March 2010)
  - Configuration, monitoring and trouble shooting of Mail Server
  - Deletion, assigning Quotas and and providing support for Nokia Siemens Networks users
  - Configuration monitoring and troubleshooting of FTP
  - Manage clients helpdesk for telephonic and mail Support

- **"Website and application development"**
  Application developer at E-Ware TECHNOLOGIES PVT LTD Gurgaon, India (October 2008 - December 2008)
  - Worked on ASP.NET infrastructure and learned in's and out's of visual studio
  - Obtained familiarity with common language run-time and just-in-time compilation
  - Worked on MySQL on backend

Publications

- Learning to Recommend Diverse Items over Implicit Feedback on PANDOR with Charlotte Laclau, Massih-Reza Amini, Accepted in RecSys 2018
- Health Monitoring on Social Media over Time in IEEE Transactions on Knowledge and Data Engineering Journal with Sihem Amer-Yahia, Marianne Clausel, Majdeddine Rebai, Mai Thai Son and Massih-Reza Amini, January 18, 2018
  Publication URL: `https://ieeexplore.ieee.org/document/8263414/`
- KASANDR: A Large-Scale Dataset with Implicit Feedback for Recommendation in SIGIR 2017, TOKYO, JAPAN with Charlotte Laclau, Massih-Reza Amini, Gilles Vandelle and Andre Bois-Crettez
  Publication URL: `https://dl.acm.org/citation.cfm?id=3080713`
- Health Monitoring on Social Media over Time in SIGIR 2016, PISA, Italy with Shashwat Mishra, Sihem Amer-Yahia, Marianne Clausel, Massih-Reza Amini
  Publication URL: `https://dl.acm.org/citation.cfm?doid=2911451.2914697`
- Representation Learning and Pairwise Ranking for Implicit Feedback in Recommendation Systems with Mikhail Trofimiov, Oleg Horodnitskii, Charlotte Laclau, Yury Maximov, Massih-Reza Amini
  Publication URL: `https://arxiv.org/abs/1705.00105`

Technical Skills

- **Github** `https://github.com/sumitsidana`
- **General purpose Programming Languages** - Java, C++, Python, Scala
- **Distributed Frameworks** - SPARK
- **Deep Learning Framework** - Tensorflow
- **Scripting Languages** - Shell script, Bash Commands
- **Web based programming languages** - HTML
- **Databases** - MySQL, PostgreSQL, SQLite
- **Text Editors** - Vim, Gedit, MSOffice
- **Operating Systems** - Worked on Linux (Debian, Ubuntu), Windows 7/8

Other Activities

- **Competitive Programming**
  - An active participant of programming events in the college
  - Done a course on Competetive programming (six months)
  - Programming Profiles:
    - Codechef - profile: sumitsidana
    - SPOJ - profile: Sumit Sidana
    - Division 2 in Codeforces - profile - sumit17
    - Kaggle - profile - Sumit Sidana

Education

Université Grenoble Alpes (UGA), Grenoble, France
- *Doctor of Philosophy (Ph.D.) in Recommender Systems and Online Advertising* **2015 – 2018**

International Institute of Information Technology (IIIT), Hyderabad, India
- *Master's of Technology in Computer Science and Engineering* **2011 – 2013**
- Cumulative Performance Index (CPI) of **7.71** (on a scale of 10)

Swami Vivekanand Institute of Engineering and Technology, Punjab, India
- *Bachelor's of Technology in Information Technology* **2004 – 2008**
- Cumulative Percentage of **78.4** (on a scale of 100)

Scholastic Achievements

- Secured **All India Rank 368** (among 224,160 students) in GATE 2013.
- Secured **All India Rank 1146** (among 136,027 students) in GATE 2011.
- Secured **315/340** in GRE General Test given on 11/11/2013

Course Projects

- **"Distributed SQL Query Optimization"**
  Advances in database systems under mentorship of Dr. Kamal Karlapalem, CSE, IIIT-H (February 2013 - April 2013)
  - Worked in a team of 2
  - Designed a cost-based query optimization module between three sites
  - Designed and implemented the SQL query parser and executer in Java.
- **"Opinion Mining in JAVA"**
  NLP Applications under mentor: Dr. Prashanth Mannem, CSE, IIIT-H (March 2013 - April 2013)
  - Built a supervised sentiment analysis system for movie reviews and news articles in a team of 2.
  - Dataset: `https://www.cs.cornell.edu/people/pabo/movie-review-data/review_polarity.tar.gz`
  - Feature engineering - N-grams, Parts of Speech Tagging, Stop-Word removal etc.
  - Term paper for the project is available at: `http://1drv.ms/1BTaYh5`
- **"Face recognition system"**
  Pattern Recognition under Prof. Anoop M. Namboodiri, CSE, IIIT-H (October 2012 - December 2012)
  - Implemented the paper - `http://www.face-rec.org/algorithms/PCA/jcn.pdf`
  - Data: Training set size of 16 images
  - Extracted useful features from the faces using Principal Component Analysis
  - All the images were projected on reduced dimension set (given by PCA) and then testing was performed on new and distorted images
- **"Insights in accidents"**
  Data Mining under Dr. Kamal Karlapalem, CSE, IIIT-H (October 2012 - December 2012)
  - Various features were learned affecting accidents such as light conditions, weather conditions, gender, age group, fatal accidents vs. mild accidents
  - Learned to use Weka - a popular data mining tool
  - Wrote a term paper on obtaining insights in accidents
- **"A Java-based Search Engine"**
  Information retrieval under Prof. Vasudeva Verma, CSE, IIIT-H(January 2012 - April 2012)
  - Dataset of 23 GB Wikipedia articles.
  - Built an inverted index (word to document mapping) to process, parse and give fast query results via TF-IDF page ranking.
  - Implemented classic Google Page Rank algorithm over the top of this search engine
- **"Modelling to solve Sudoku Constraint Satisfaction Problem"**
  Optimization Methods under Dr. C.V. Jawahar, CSE, IIIT-H (January 2012 - April 2012)
  - Modelled and solved Sudoku as a Constraint Satisfaction Problem in MATLAB

Positions of Responsibility

- **Paper reviewer, Knowledge Based Systems** (January 2017)
- **Teaching Assistant, Discrete Mathematics** (January 2012 - April 2012)
  - Managed the course along with the instructor Dr. Ashok Kumar Das by teaching and conducting examinations with an enrollment of more than 100 students
  - Topics included Number theory, generating functions, permuation and combinations, Complexity theory.
- **Teaching Assistant, Principles of information security** (July 2012 - December 2012)
  - Conducted a course on Principles of information security with an enrollment of more than 150 students along with Dr. Ashok Kumar Das, CSE, IIIT-H
  - Topics included discussion on cryptography, digital signatures, quantum cryptography

Synopsis

- An effective communicator in English with excellent relationship building and interpersonal skills. Strong analytical problem solving and organizational abilities.
- A data scientist having been involved in multiple projects which helped gaining ample experience in working with large datasets and large scale learning.
- An efficient coder with proficiency in both algorithmic and object-oriented coding with focus on always writing optimized and scalable code
- A Mathematician with research interests in Linear Algebra, Graph Theory, Probability and Statistics.
- A person who possesses a flexible and detail oriented attitude