

# Tuhin Sharma

<https://www.tuhinsharma.com>

tuhinsharma121@gmail.com

+91-8861557799

## Career Profile

- 10+ years of core analytics/data science experience in helping businesses leverage data.
- Expertise in Machine Learning, Deep Learning and Natural Language Processing.
- Expertise in designing/building/deploying end-to-end Machine Learning Systems in B2B and B2C setup.

## Technical Skills

- Languages : Python, Scala
- Tools : Spark, Flink, Kafka, Mongo, Druid, PostgreSQL, Pinot, Snowflake, Quadrant, Milvus, Prophet, Sklearn, Mxnet, Pytorch, Docker, Fastapi, Mlflow, Superset.

## Work Experience

Senior Principal Data Scientist, Red Hat, Bangalore, India	Jul 2021 – Present
Lead AI Architect, Hypersonix, Bangalore, India	Nov 2020 – Jun 2021
Co-Founder and CEO, Binaize, Austin, USA	Dec 2018 – Nov 2020
Senior Data Scientist, Red Hat, Bangalore, India	Feb 2017 – Sep 2017
Data Scientist, IBM Watson, Bangalore, India	July 2014 – Feb 2017

## Education

**Indian Institute of Technology Roorkee (IITR)** **2012-2014**

M.Tech in Computer Science and Engineering, Data Mining Lab with CGPA 9.29  
*Thesis:* A Generalized Framework for Quantifying Trust of Social Media Text Documents

**Indian Institute of Engineering Science and Technology Shibpur (IESTS)** **2008-2012**

B.E. in Computer Science and Technology with aggregate 80% marks  
*Thesis:* Dynamic Network Traffic Data Classification for Intrusion Detection Using Genetic Algorithm

## Recent/Relevant Projects Details

- ❖ **JarvisAI (RedHat)** is a suite of GenAI products that focus on increasing operational efficiency at organizational level across various business units. Architected the end to end workflow leveraging **QDRANT**, **OPENSIFT AI**, **ROSA** p4 clusters.
- ❖ **Telesense (RedHat)** is a suite of analytics products that focus on transforming customer usage data into specific, actionable insights.
- ❖ **Hawkeye (Hypersonix)** helps enterprises drive profitable revenue growth by providing a system of actionable intelligence. Designed the end to end architecture of **ML Pipeline** on **AWS EMR**, **Mlflow**, **Apache Superset** and established TDD for building ML solutions.
- ❖ **Binaize (Binaize)** helps small and medium e-commerce businesses increase online conversion with the same marketing budget. Built, deployed and sold an intelligent and efficient version of Multi-Armed Bandit (**MAB**) platform to **Shopify** Customers.
- ❖ **Fabric-8-analytics (RedHat)** provides recommendations for the manifest file of a developer in terms of alternate package, companion package and outlier package. Built and deployed Bayesian Network (**BN**)

based **Recommender System** for **Eclipse Che**. Code is available at <https://github.com/fabric8-analytics/fabric8-analytics-stack-analysis>

- ❖ **Demand Insight (IBM Watson)** provides insights about demands of products (SKUs) to be sold based on weather and social information about respective store locations in retail domain. Built Attribute based Forecasting model (ABF) using **PYSPARK** for **Meijer** across **200** locations in the USA.
- ❖ **Predictive Customer Analytics (IBM Watson)** provides insights about Churn, CLV, Segmentation etc. in the customer domain. Built Tree based models (RF, GBT) using **PYSPARK** for Big Data Data platform on **Bluemix**.
- ❖ **Social Media Analytics (IBM Watson)** provides insights about most talked about incidents about financial institutions (banks, insurance etc.) by interpreting twitter data. The insights could be used for targeting potential customers, improving customer service and even discovering potential brand ambassadors. Built and deployed Event Detection Models (EDM) for banks like **ICICI, KOTAK MAHINDRA** and **YESBANK**.
- ❖ **FreeTime Designer (IBM Watson)** app for **SXSW 2015**. The conference had more than **2000 music events** and **250 film events**. Based on a person's interests, the app will come up with an automated schedule to maximize the user satisfaction. Close to **40,000 people** used the app during the conference. Built and deployed DP based scheduling algorithm **Weighted Activity Selection**.
- ❖ **Score With Data (IBM Watson)** provides game highlights for cricket world cup 2015, solely based on the twitter data. Built AI system to crawl tweet feeds during the match, clean and cluster them into stories and eventually generate crisp and concise highlights. From February 14 to March 29, ingested and analyzed around **1M tweets** per match on an average, reaching a peak of **21M** during the semi-final and final matches. The results were broadcasted on **CNN-IBN** everyday, after individual match was concluded during whole season of **CWC2015**. <https://www.ibm.com/blogs/research/2015/04/cricket-big-data/>

## Recent Mentoring/Consulting (Not Exhaustive)

**Threat Hunt for Cyber Security (AT&T)**  
2019

Jan 2019 – Mar

Set up the cloud platform for data pipeline to process 20B records everyday to detect fraud in real-time.

**Event Management for Coding Platform (WhiteHat Jr)**

Dec 2018 – Jan 2019

Architected the event management platform to capture clickstream data during Mentor-Student interaction. It was used to improve the Students' learning experience and better training for the Mentors. The company got an exit of \$300M.

**Recommendation System (Republic TV)**

Jul 2018 – Dec 2018

Built and deployed Probabilistic Graphical Model (PGM) based **Recommender System** for **Republic TV** with a response time of ~**50 ms**.

**Movie Recommendation System (Fastfilmz)**

Nov 2017 – Feb 2018

Built and deployed Hybrid (Content based and Collaborative filtering based) Recommendation system for suggesting South Indian regional movies. The company got an exit of \$11M.

## Publications

1. **Tuhin Sharma**, Durga Toshniwal (2016), "**A Generalized Framework for Quantifying Trust of Social Media Text Documents**", Proceedings of **12th International Conference on Machine Learning and Data Mining (MLDM 2016)**, published in LNCS 9729, pp. 692-713, by Springer-Verlag Berlin Heidelberg, 2016.
2. **Tuhin Sharma**, Durga Toshniwal (2014), "**A Generalized Relationship Mining Method for Social Media Text Data**", Proceedings of **10th International Conference on Machine Learning and Data Mining (MLDM 2014)**, published in LNCS 8556, pp. 376-392, by Springer-Verlag Berlin Heidelberg, 2014.
3. Sahisnu Mazumder, **Tuhin Sharma**, Rahul Mitra, Nandita Sengupta and Jaya Sil (2012), "**Generation of Sufficient Cut Points to Discretize Network Traffic Data Sets**", Proceedings of Third International

Conference on **Swarm, Evolutionary, and Memetic Computing (SEMCCO)** published in LNCS 7677, pp. 528-539, by Springer-Verlag Berlin Heidelberg, 2012.

4. Rahul Mitra, Sahisnu Mazumder, **Tuhin Sharma**, Nandita Sengupta and Jaya Sil (2012), “*Dynamic Network Traffic Data Classification for Intrusion Detection using Genetic Algorithm*”, Proceedings of Third International Conference on **Swarm, Evolutionary, and Memetic Computing (SEMCCO)**, published in LNCS 7677, pp. 509-518, by Springer-Verlag Berlin Heidelberg, 2012.

## Patents

1. **Tuhin Sharma**, Krishna Kummamuru, Ravi Tejwani, Jeff Eisen, “[User experience using social and financial information](#)”
2. **Tuhin Sharma**, Bharath Ganesh, “[Use of asset and enterprise data to predict asset personality attributes](#)”
3. Dhandapani Shanmugan, Bharath Ganesh, **Tuhin Sharma**, Jothi Subramani, “[Cognitively inducing Sleep Cycles by leveraging Wearables](#)”
4. Dhandapani Shanmugan, Bharath Ganesh, **Tuhin Sharma**, Gerin Jacob, “[Mitigation Of Perishable Product Waste Through Promotion Recommendation](#)”
5. Dhandapani Shanmugan, Bharath Ganesh, **Tuhin Sharma**, Jothi Subramani, “[Enabling Wearables To Cognitively Alter Notifications And Improve Sleep Cycles](#)”

## Other Interests/Achievements

- **GIDS Speaker (2024)** - Presented a talk about “[Navigating Innovation with Open Hybrid Cloud and Openshift AI](#)”
- **GIDS Speaker (2023)** - Presented a talk about “[Pybandit: A Website Optimization Framework for E-commerce SMBs](#)”
- **ODSC Europe Speaker (2022)** - Presented a talk about “[Eagleeye: Data Pipeline for Anomaly Detection in Cyber Security](#)”
- **Devconf.cz Speaker (2022)** - Presented a talk about “[Data Pipeline for Anomaly Detection](#)”
- **GIDS Speaker (2021)** – Presented a talk about “[Signature Verification in Banks using Few Shot Learning](#)”
- **Cohort Member at Techstars Austin (2020)** – Got selected (10 out of 1189 companies worldwide) in prestigious Techstars Austin startup accelerator. Raised \$120k capital.
- **Pybandit (2019)** – Started FOSS initiative to build scikit-learn for AB-Testing which is first of its kind.
- **Datahack Summit Speaker (2019)** – Presented a talk about “[Federated Learning using Deep Learning](#)”
- **O’reilly AI Conference London Speaker (2019)** – Presented a talk about “[Anomaly detection in smart buildings using federated learning](#)”
- **ODSC India Speaker (2018)** – Presented a talk about “[Hybrid Recommendation Systems in News Media using Probabilistic Graphical Models](#)”
- **O’Reilly Author (2017)** – Deep Learning with MXNet.
- Won **First Runner Up at Techathon (2015)** (IBM Watson) for developing an app which creates automatic highlights of a cricket match using events mined from social media.
- **Excellence and Eminence Award (2015)** (IBM Watson) - Critical and Timely contribution to content building on Next-Gen Platform.
- **Goyal Award (2014)** (IIT Roorkee) Topper of the batch of 2012-2014 in M.Tech (CSE) programme.