Shwet K Mishra

Data Scientist | ML Engineer

📞 (+91) 7376-344-342 | ✉ shwetknit@gmail.com | ⌨ shwetkm | in shwetkm

Experience

DataWeave   *Bengaluru*
Data Scientist   *March 2019 - PRESENT*

- Created a **Retail Knowledge Graph** to store dark unstructured data from web in a structured form to a Graph DB(*Neo4j*). Used this Knowledge Graph to build an **Attribute Tagging** module which could tag attribute-values from e-commerce product descriptions.
- Implemented a research paper **"Multi-Level E-Commerce Product Categorization Via Machine Translation"** to build a solution for Product Categorization.Task was to identify appropriate product taxonomy (e.g. Footwear > Men > Formal > Oxford) for any retail product(Footwear and Clothing category) from given title and description. Here, a Seq2Seq approach is used for hierarchical classification.
- Trained retail specific **Fasttext Word Embeddings** and used them for text similarity to improve a Product Matching engine where only **Tf-Idf** was being used earlier. Used Apache Solr for text indexing and search to reduce candidate sets for product matching.

Exzeo Softwares   *Noida*
Data Scientist   *October 2017 - March 2019*

- Built **Image classification** tool to predict *Roof Shape* out of Flat, Gable and Hip from their **satellite images**. Implemented techniques like *Cyclic learning rates, Test Time Augmentation,* etc. to boost the performance of image classifier.Built **Delivery and Training pipelines** using *Luigi* framework to automate the entire process.
- Built a **Risk Modelling** system to understand risk associated with insured/non-insured properties.Dealt with *class imbalance* by majority class under-sampling followed by stacking to produce results.
- Planned and developed solution for a project to *extract domain specific information* from variety of insurance documents. Used tesseract to convert pdf images to text. Devised methods with combination of heuristics and learning models to extract insurance specific Key-Values from reports and mail attachments. Built **API** to interact with **Mailgun** mailing service to deliver results.
- Built a product for insurance underwriters to understand **Claim Cost** w.r.t. *Claim type, Property features and Complaint Call Transcripts*. Trained Claim Cost prediction model using TF-Idf vectors and other existing features to predict the cost bracket for a filed claim. Built **Claim Cost Exploration Visualisation Tool** to visualise properties on *Florida Map*.

Tech Mahindra Growth Factories Ltd.   *Hyderabad*
Data Scientist   *August 2016 - October 2017*

- Worked as a subject matter expert for e-learning venture of Tech Mahindra- UpX Academy. Worked on several Case Studies involving *Machine Learning, Hadoop, Spark and NoSQL Databases.*
- **Optimised marketing campaigns** and cut costs by **25%** from useful insights through Data Analysis and Machine learning models.Built a Flask REST API to identify prospective customers based on their data in **CRM** tool

Education

Kamla Nehru Institute of Technology   *Sultanpur, U.P., India*
Bachelor of Technology in Computer Science & Engineering, Percentage: 78.54   *May 2016*

MOOCs
- **Deep Learning Specialization**, Deeplearning.ai on Coursera
- **Machine Learning by Andrew Ng**, Stanford Univerity on Coursera
- **Introduction to Big Data**, University of California San Diego on Coursera

Skills

Languages	Python, R, Cypher, MySQL
Frameworks and Tools	Sklearn, PyTorch, Fastai, Kafka, MongoDB, AWS, Docker, Neo4j, NetworkX, Flask, Spark, Solr, Luigi, RabbitMQ, Redis

Additional Experiences

Kreate Hackathon by Digital Ocean

- **Won 1st prize** in *Kreate Hackathon at Kalaari Capital organized by Skillenza & DigitalOcean.*
- Built an **Image Spam Buster** for Mall91 where we had to identify spams like promotional text image, vulgar images, etc. from the user generated contents. We used multi-pronged approach which involved ensemble of 3 models, i.e. Text Classification Model for English, Hindi and Mixed content, Heuristics Model based on known "bad words" and Image Classification Model for non-text images.
- Got featured on a **YourStory** article for winning the Hackathon.

Brainwaves 2019 by Societe Generale

- **Secured 7th Rank**(out of 10,200+ participants) in *Brainwaves 2019 Machine Learning Hackathon organized by Societe Generale* on Hackerearth.
- Built a model to predict the outcome of a customer complaint based on textual content and other features of the complaint.

Innoplexus Machine Learning Hackathon

- **Secured 4th Rank**(out of 1961 participants) in *Innoplexus Machine Learning Hackathon organized by Innoplexus on Analytics Vidhya*.
- Build a model to predict category of articles published on news websites.

Tech Talks

- Gave a talk on **Advancements in Word Embeddings: From Word2Vec to BERT** at DataWeave
- Speaker at **PyData** Delhi Meetup; Delivered a presentation on **'How to build Data Science Pipelines'**.