

Mohit Mayank

Lead Data Scientist @ Outplay (mohitmavank.com)

- Helping SaaS startups setup and scale their AI game 🚀
- 7+ years of industrial experience in Sales, Finance, Retail and other domains.
- Projects in Generative AI, NLP, CV, Audio, Graphs, RL, DL and ML.
- Guest lecturer and collaborator at multiple Universities.
- Author of "[Lazy Data Science Guide](#)" & Creator of "[Jaal](#)" (open-source)
- Winner of [TCS SUPERCoder 2019](#) - a company-wide coding competition.



WORK EXPERIENCE

Lead Data Scientist

Outplay

Nov 2021 - Present

Remote

Generative AI (Text & Audio)

– **SalesGPT**: Training and Deploying inhouse LLMs to make it more domain aware for Sales related downstream tasks. Explored 3rd party models like GPT-4 for evaluation. Required extensive prompt engineering and quantization.

Keywords: GPT4; ChatGPT; LLaMA; Quantization; QLoRA; Flash Attention; LangChain

– **Inhouse Text-to-Speech**: Trained Inhouse TTS model to generate high quality and fast audio for provided text prompt. Able to control pitch, speed, etc using SSML.

Keywords: VITS; Text Analysis; Acoustic Model; Vocoder;

– **Voice Cloning**: Inhouse models to support creation of fast and accurate voice clones of Non-Western voices with very few examples or with fine-tuning approach.

Keywords: XTTSv2; tortoise-tts; Transformers; Quantization; LoRA

Products

– **PhoneVerify.ai**: AI powered phone number validating product that 10x the dial-to-connect ratio. Automatically calls and converses with prospect to identify if it has reached correct number or not, find the best to call, detect call outcome, etc.

Keywords: Speech Recognition, Text-to-Speech, LLM, Dialog Management, Twilio

– **Conversation Intelligence**: AI powered meeting analysis product that joins ongoing calls to identify prospect's concerns, suggesting answers, perform live mood detection and after the call ends, performs further advanced analysis!

Keywords: Speech Recognition, Speaker Identification, Mentions Identification, etc.

– **Email Assistant**: Multi-task AI assistant that, (1) writes a highly personalized mail from scratch for a sales representative, and (2) provides recommendations to enhance existing sales mails by gamifying the complete process.

Keywords: Mail Scoring, Sentiment, Tones, Spam words, Paraphrasing, etc

Audio Intelligence

– **In-House Transcription System**: speech-to-text (ASR) services for live and offline sales call transcription; comparable to the likes of AWS Transcribe.

Keywords: Wav2Vec2; HuBERT; WordBoost; Custom Vocabulary; Websocket

– **Speaker Diarization and Identification**: Segregate and recognition different speakers in a call by comparing against the available speaker's voice prints

Keywords: SpeechBrain; Bi-encoder; Audio Embeddings

Senior Data Scientist

AlgoAnalytics

March 2021 - Oct 2021

Pune, India

Natural Language Processing (NLP)

– **Indian Companies Knowledge Graph**: Create a KG containing the enterprise-level details for Indian companies (like Crunchbase); analyze graph to find similar company and recommend potential acquisitions and investments

Keywords: Knowledge Graphs, Recommendation System, Information Extraction, etc

– **Social Media Trend Analyzer**: Analyze the trend of product, person or any entity on Social media like Twitter.

Keywords: NER, Social media scraper, Trend forecasting, Sentiment, Tones, etc

– **Financial Events Extraction**: extracting relevant financial real-world events like acquisitions or investment, from unstructured texts like news, social media posts, etc.

Keywords: Entity Linking; NER; Website scraping

– **Stock Market Analysis**: identify hidden signals on price, volume and news related to stock market by leveraging NLP algorithms; Leverage the data to make market predictions.

Keywords: News Sentiment Analysis, Classification, Timeseries forecasting, etc

EDUCATION

Bachelor of Engineering (B.E.), Electronics and Telecommunication

Savitribai Phule Pune University
(formally Pune University)

2012-2016

Pune, India

ACHIEVEMENTS

Rank 1 – TCS SUPERCoder 2019

Beat >20k associates to win the prestigious TCS wide competitive coding competition

Winner – Digitate Ideathon 2019

Topic: [Log analyzer](#) – Detect events, find anomalies and perform root cause analysis

Runner up – Digitate Ideathon 2018

Topic: [Interactive KG learning interface](#) – asks relevant questions to build knowledge graphs

“On the spot award”

For excellence in quick solution proposal and fast deployment

“Star team award”

Collaborative effort for problem management

ARTICLES

- Building chatbot with Rasa ([link](#))
- Introduction to Genetic Algorithm ([link](#))
- Reinforcement Learning with Multi-Arm bandit ([link](#))
- Generating images using text: The dawn of the AI dreamers ([link](#))
- A Guide to Python Good Practices ([link](#))
- A Guide to Knowledge Graphs ([link](#))
- Guide to fine-tuning Text Generation models: GPT-2, GPT-Neo and T5 ([link](#))
- Explainable AI: Language Models ([link](#))

PUBLICATIONS

- [DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection](#) (A deep neural network based framework to detect fake news from short text by leveraging open domain context using knowledge graphs)
- [Intrinsic analysis for dual word embedding space models](#) (Comparative analysis of a total of 84 different word embedding models)

Data Scientist

Tata Research Development and Design Centre, TCS

Aug 2016 - Feb 2021

Pune, India

Projects: *Digitate and Intelligent Services Network*

Natural Language Processing (NLP)

– **Question-Answer chatbot**; virtual assistant for project's flagship solution that increased daily impressions by 10% and reduced human intervention by 40%

Keywords: Chatbot, Question and Answer identification

– **Issue extraction**; identify different types of system alerts from IT ticket's descriptions for descriptive analysis of enterprise

Keywords: Topic Identification, Text Similarity

– **Resolution notes mining**; mine L1/L2/L3's alert resolution notes to generate rules to automate frequently occurring errors

Keyword: Pattern finding and matching, Association rule mining

Computer Vision (CV)

– **Business Card Reader**; OCR engine to extract entities from multi-lingual business cards for ultra-fast personnel contact onboarding

Keywords: Optical Character Recognition, NER, Validity check

– **Math formula Recognition**; extract handwritten math expressions from documents for fast digitalization and education purposes

Keywords: Optical Character Recognition, LaTeX n-gram model

Forecasting

– **Marathon prediction**; create real-time model to predict NY marathon runner's race completion time with over 80% accuracy

Keywords: Sequence Prediction, LSTM; Clustering

– **Enterprise revenue prediction**; revenue forecasting of company's hierarchical business units with ~91% accuracy

Keywords: Time series forecasting; ARIMA, Holt Winters, HTS

Optimization

– **System expert assignment**; given an IT system alert, automatically find the best L1/L2/L3 resolver and assign the task; done by performing multiple objective optimization, leading to reduction of overall MTTR by 30%

Keywords: Ranking algorithms; Multi-objective optimization; Resource allocation

– **Personalized pricing model**; find optimal Batch analytics pricing by minimizing jobs in analysis with minimum adverse effect on SLA prediction accuracy

Keywords: Graph reduction; Subset finding; Price optimization

Causality Analysis

– **Root-Cause analysis**; incident management on structural graph of the enterprise to identify correct root cause for an system level event.

Keywords: Correlation vs Causality, Graph traversal, Multi-modal analysis

– **What-if analysis**; model causality graph of an enterprise to perform intervention on behavior of services and ask counterfactual business analysis questions

Keywords: Product Behavior Simulation, Confounding analysis, Prediction

Recommendation System

– **Enterprise Sales**; hybrid recommendation system to generate B2B suggestions for company's Sales team with major focus on accuracy and explainability

Keywords: Recommendation Systems, Cross-Sell, Up-Sell, Day 0 Sell

Research Intern

Centre for Development in Advanced Computing


2015


Pune, India

Speech Recognition

Design and developed IoT **Speech Recognition engine** in Indian languages and deployed it on low-power hardware as a low-cost solution for the rural population.

Contacts

 mohitmayank1@gmail.com

 +91-7350272674

 Pune, India

 mohitmayank.com

 imohitmayank

LECTURES/TALK

- [Lecture on Knowledge Graph - Part 1](#) (Introduction, Ontology, Query, Creating KG and KG Embeddings)
- [Lecture on Knowledge Graph - Part 2](#) (AI in KG, KG tasks and applications)
- [The Art behind Fractals](#) (An interactive talk I gave on the beauty behind the Fractals)

BOOK/NEWSLETTER

- [A Lazy Data Science Guide](#) (A guide book on data science for busy and equally lazy Data Scientists 😊)
- [The ML Dojo](#) (the single place to get daily updates on the latest in the field of Artificial Intelligence 🧠 and Machine Learning 🤖)

PERSONAL PROJECTS



- [Sankshep](#) (Sankshep is a free website that uses ChatGPT powered AI assistant to provides summaries and a way to chat with your research papers)
- [Jaal](#) (Python package with more than 150 stars on Github. It is an interactive network visualization tool built using Dash and Visdcc)
- [Link Summarization](#) (a web demo that can summarize a website. Takes a website link to parse the text and summarize the content)
- [Law Finder](#) (AI model to find the most suitable Indian Penal Code law if provided with a crime's description)
- [Emoji Finder](#) (AI model that suggests the suitable emoji based on the input sentence)
- [Q learning Demo](#) (An interactive playground to experiment and learn the Reinforcement Learning topic - Q Learning)

INTERESTS

Books/Movies

Gaming

Fitness

Competitive Coding

Technical Writing

Guitar

Scan QR code!

