

SENTHIL KUMAR

senthilkumar.m1901@gmail.com • +91-984-186-9609 • [GitHub](#) • [LinkedIn](#) • [Blog](#) • Chennai, India

- A Data Scientist
 - specialized in building production-grade NLP applications utilizing ML and DL techniques
 - capable of co-developing ML applications by following robust coding practices
 - who has extensively used state-of-the-art Transfer learning models
- An aspiring developer who strives to build clean, modular software applications.
- Total Exp: 12 years; Data Science Exp: 8/12 years

EXPERIENCE

Lead Data Scientist, Ford, Analytics Team

May 2018 – Jun 2022

- A hands-on **data science developer** who co-developed ML applications with in-company NLP experts from the US
- Worked for teams such as AI Advancement Center, Customer Experience and Manufacturing Operations
- I contributed in the end-to-end ML application development
 - from data acquisition, cleaning, labeling and preprocessing,
 - to model development, deployment and maintenance
- A **Python Trainer** and **Technical Interviewer** of NLP candidates across analytics teams

Assistant Manager, LatentView Analytics (LV)

Apr 2014 – Apr 2018

- Roles I played: **Data Scientist, Project Delivery Manager**
- Utilized Python, SQL, ML and NLP Skills to uncover answers from social media text data
- Accountable for the delivery of Social Media Analytics projects of 8+ members
- Responsible for project scoping, resource planning and work distribution
- LinkedIn Recommendations:
 - *"... extraordinary dedication contributed significantly to growing our analytic practice..."* - Customer of LV
 - *"....Sincere, driven, articulate and utterly committed ..."* - Skip-level Reporting Manager at LatentView

Lead Analyst, Beroe Inc

Jul 2010 – Dec 2013

- Produce Market Research reports on how to procure indirect spend categories
- *"... well organized, innovative ... and always ready to go the extra mile"* - Client Engagement Manager

EDUCATION & CERTIFICATIONS

B.E - Electronics - 8.6 CGPA

Madras Institute of Technology,
'06 - '10

12th Grade - 95% | 10th Grade - 92%

State Topper in Physical Science paper,
'06 TN Engineering Entrance Exam

DeepLearning.ai Specialization

Coursera (5 courses), Dec'18 - May'19

GCP Big Data & ML Fundamentals

Google - Coursera, Apr'21

Applied ML and Applied Text Mining

Michigan Univ - Coursera (2 courses), Dec'17 - Jan'18

SQL (GCP BigQuery) Fundamentals

Kaggle Learn, Feb'22

ML Fundamentals for Structured Data

Kaggle Learn (2 courses), Jan-Feb'22

Probability and Statistics Fundamentals

LinkedIn Learning (2 courses), Dec'21

TECHNICAL SKILLS

Languages

Python, SQL(basics),
Markdown, Linux Shell (basics)

Tools

Git, WSL, Docker, Kubernetes,
Poetry (Python env), Conda, PyCharm

Python Libraries (extensive usage)

Pandas, SpaCy, Re (Regular Expressions),
Transformers, Sklearn, PyTorch

Python Libraries (working knowledge)

PySpark, FastAPI (REST API),
Streamlit (UI), Altair (viz)

KEY PROJECTS

BERT Fine-tuned Aspect-based Sentiment Analysis Pipeline

[More Details](#)

- Built a reusable Sequence Classification ML Pipeline which converts customer comments into trackable Aspect and Sentiment pairs
- The ML Pipeline used BERT-fine-tuning and it helped yielding 85%+ F1 score with minimal annotated data for more than 25+ classes
- Incorporated easy to use human-in-the-loop annotation and model monitoring scripts

Personally Identifiable Information (PII) Detection using Named Entity Recognition (NER)

[More Details](#)

- Anonymized PII in text data by building a NER system using **RoBERTa Fine-tuned Transformer model**
- Bootstrapped the training data using Spacy rules (thus easing the annotation process by not starting labeling from scratch)
- Deployed an asynchronous inference REST API (using FastAPI and K8s) that can be plugged into multiple applications

NLP Semantic Search Pipeline

[More Details](#)

- Created a “digital thread” by connecting two automotive domain specific data sources
- The data sources contain technician comments about issues before the launch of a vehicle
- The digital thread was established by assigning NLP-based semantically matching common part descriptions in comments in both datasources
- Purpose: Predict the prominent issues about to occur in a downstream data source by reviewing the issues much earlier in the launch cycle

Reusable Text Data Clustering Pipeline

[More Details](#)

- Built reusable Text Clustering pipeline with simpler Python APIs for non-NLP analysts

PROMOTIONS & AWARDS

Ford: Senior Data Scientist to Lead Data Scientist Promotion

In Nov'2019, after 1.5 years of joining Ford

Ford: Asia-Pacific Recognition Award

*Won in May '19
for successful spearheading of a project*

LatentView: Senior Analyst to Assistant Manager Promotion

In Oct'16, after 2.5 years of joining LatentView

LatentView: Encore Award

*Won for company-wide best performance
for the Jul-Sep 2016 quarter*

Beroe: Promoted twice in my first company

During my 3.5 year stint in Beroe

Beroe: Knowledge Contributor Awards

*Won twice for company-wide
best performance in Q1 and Q2 calendar year
2013*