

# SENTHIL KUMAR

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

---

- A Data Scientist
  - specialized in building production-grade NLP applications utilizing ML and DL techniques
  - capable of end-to-end ML project execution
  - capable of translating business needs into analytical requirements
- 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 – Present

- Play the role of a hands-on **data science developer** who employs state-of-the-art ML and DL techniques for NLP Applications
  - for analytics teams such as *AI Advancement Center*, *Customer Experience* and *Manufacturing Operations*
- Contribute in the end-to-end ML application development
  - from data acquisition, cleaning, labeling and preprocessing,
  - to model development, deployment and maintenance
- **Python Trainer** | **Technical Interviewer** of NLP candidates across analytics teams

### Assistant Manager, LatentView Analytics

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 Recommendation: "... *extraordinary dedication contributed significantly to growing our analytic practice...*" - F100 Tech Client stakeholder
- LinkedIn Recommendation: "....*Sincere, driven, articulate and utterly committed* ..." - Skip-level Reporting Manager at LatentView

### Senior Consultant, Capgemini, Spend Analytics

Jan 2014 – Apr 2014

### Lead Analyst, Beroe Inc

Jul 2010 – Dec 2013

- Produce Market Research reports on how to procure indirect spend categories

## 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 2021

### **Applied ML and Applied Text Mining**

Michigan Univ - Coursera, Jan '18

### **SQL (GCP BigQuery) Fundamentals**

Kaggle Learn, Feb 2021

### **ML Fundamentals for Structured Data**

Kaggle Learn (2 courses), Feb 2021

### **Probability and Statistics Fundamentals**

LinkedIn Learning, Dec 2021

## 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 scripts 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 both **RoBERTa Fine-tuned Transformer model** and Rules-based logic
- Bootstrapped the training data using Spacy rules (thus easing the annotation process by not starting labeling from scratch)
- Achieved an F1 score of 89% for detecting the PII entities
- Deployed an asynchronous Inference REST API (using FastAPI and K8s) that can be plugged into multiple applications

### NLP Semantic Search Pipeline

[More Details](#)

- Created “digital threads” for connecting automotive data sources
  - which has technician comments about issues before the launch of a vehicle,
  - by assigning semantically matching common part categories to every issue in both data sources

### 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*