

MANU JOSEPH

4814 Prestige Kew Gardens, Yemalur Road, Bangalore, Karnataka

+91-9552668639 manujosephv@gmail.com linkedin.com/in/manujosephv github.com/manujosephv
Semantic Scholar(2086742581) deep-and-shallow.com

Summary

Manu Joseph is self-made data scientist with 10+ years of cross-functional experience spanning across Analytics, Software Engineering, and Supply Chain consulting. He has extensive experience working with many Fortune 500 companies in enabling digital and AI transformations in their processes, specifically in Machine Learning based Demand Forecasting. He is a passionate and pragmatic researcher focusing on time series, tabular data, and NLP, with learning paradigms like self-supervised learning & transfer learning. He also developed and maintains an open-source library - PyTorch Tabular (750+ stars) - which makes deep learning for tabular data easy and accessible.

Education

Symbiosis Institute of Business Management, Bangalore

2011 – 2013

Master of Business Administration (Operations Management)

Bangalore, Karnataka

College of Engineering, Trivandrum

2004 – 2008

Bachelor of Technology (Industrial Engineering)

Trivandrum, Kerala

Skills

- Machine Learning
- Python
- NLP
- Applied Research
- Deep Learning
- PyTorch
- Computer Vision
- Time Series

Experience

Thoucentric

December 2021 – Present

Head of Applied Research

Bangalore, Karnataka

- Build an Applied Research Team from grounds-up and lead the research activities for the company.
- A strong, dedicated effort to productionize cutting edge technologies to add business value to customers.
- Focused research on pushing the bar with existing customer problems or breaking new ground with new problems.
- Enable Business Development with strong Point-of-views and working demos for prospective opportunities.

Thoucentric

June 2018 – December 2021

Lead Data Scientist

Bangalore, Karnataka

- End-to-end (Conception → Formulation → Execution → Stakeholder Management) of Analytics Projects with Fortune 500 clients like Unilever, Starbucks, Beiersdorf, J&J, ABInBev, GSK, Hasbro playing the following roles.
- **Analytics Consultant** to help formulate the problem and advise on solutions using Data Science.
- **Technical Lead** to manage and mentor a cross functional team and drive the project to successful closure.
- **Data Scientist** in developing models to tackle the problem at hand.

Philips Innovation Campus

December 2015 – April 2018

Analytics Consultant (Supply Chain COE)

Bangalore, Karnataka

- End-to-end management, implementation and stakeholder management of Analytics and Data Science projects in the Supply Chain and Manufacturing Space.
- Hands-on development of Tableau, Qlikview dashboards & leading a team of analysts to develop analytical solutions.
- Piloted a machine learning solution to predict a defect using upstream data.

Entercoms

April 2013 – December 2015

Senior Consultant

Pune, Maharashtra

- Bridge gaps in the Supply Chain through analysing different data streams and solutioning through simulation modelling, statistics, business process re-engineering etc.

Cognizant Technology Solutions

September 2009 – April 2011

Programmer Analyst

Coimbatore, Tamilnadu

- Translate BRD into technical specifications and code them.

Projects

Domain Adaptation based Predictive Maintenance System | *Python, PyTorch* June 2022 - Aug 2022

- Designed and implemented an encoder-decoder model with bottlenecks to learn a domain-agnostic manifold and predict Remaining Useful Life of an Engine on target dataset without any labels.

Natural Language Querying of Relational Databases | *Python, PyTorch* June 2022 - Present

- Designed a system which uses cutting-edge models like Structure-Aware Dual Graph Aggregation Network to convert a natural language question to corresponding SQL query.
- Implemented an automated visualization system on top of the results so that the retrieved results can also be visualized with ease. Designed an interactive UI which enables the user to tweak the data retrieval and visualization system to be more easy to use.

Retrieving Relevant Products from Google & E-commerce Search Terms | *Python, PyTorch* May 2022 - Present

- Designed an NLP system to retrieve relevant product titles for a particular search term to enable white space analysis.
- Custom NER model to detect different aspect in the search term like the ingredient, claims and formats and use this information in a heuristic based downstream system

Custom Quote Generator App (Blog) | *Python, PyTorch* November 2020

- Fine-tuned GPT2 model to generate original quotes with three personas.
- Hosted the model in an REST API using FastAPI
- Packaged it up in a UI with ReactJS and hosted it using KintoHub, Firebase, and MongoDB

Machine Learning based Demand Forecasting at Scale | *Python, Sci-kit Learn, pandas* 2018-2021

- Developed machine learning engines to forecast demand in the short term, considering extraneous information like planned promotions, point of sales, stock at customer locations, weather, etc. for many CPG and Retail customers in large scale implementations.
- Gradient Boosting based Global Forecasting Models with intelligent segmentation to balance between good accuracy and execution times.
- Special handling of NPIs by leveraging SKU similarity with existing SKUs.
- Attributing generated forecasts to the features using machine learning interpretability tools like SHAP.

Correcting seller uploaded e-Commerce product images | *Python, PyTorch* April 2021

- Identify the main object(s) in an image and make the background a standard white/grey color. Salient Object Segmentation using U²-Net
- Image Super-resolution to make low resolution images up to the standard resolution.

Labelling Defect Identification on High Speed Manufacturing Line | *Python, PyTorch* April 2021

- Used Paddle-OCR for a super-fast deep learning based OCR system to improve existing OCR system by getting lower false positives.

Publications

- PyTorch Tabular: A Framework for Deep Learning with Tabular Data** (2021). <https://arxiv.org/abs/2104.13638> (https://github.com/manujosephv/pytorch_tabular)
- GATE: Gated Additive Tree Ensemble for Tabular Classification and Regression** (2022). <https://arxiv.org/abs/2207.08548> (*Under Review at AAAI 2023*)
- LAMA-Net: Unsupervised Domain Adaptation via Latent Alignment and Manifold Learning for RUL Prediction** (2022). <https://arxiv.org/abs/2208.08388> (*Under Review at AISTATS 2023*)
- AskYourDB: An end-to-end system for querying and visualizing relational databases using natural language** (2022). <https://arxiv.org/abs/2210.08532> (*Under Review at PAKDD 2023*)
- Ensemble Creation via Anchored Regularization for Unsupervised Aspect Extraction** (2022). <https://arxiv.org/abs/2210.06829> (*Under Review at SAC 2023*)

Achievements

- Authored a book on Timeseries Forecasting called **Modern Time Series Forecasting with Python** by Packt (Amazon)
- Enhanced the language model section in **NLTK** by adding many statistical language models like **Kneser-Ney**, **Lidstone**, **Stupid Backoff**, etc.
- Presented a session about **PyTorch Tabular** in an international conference - **TDC Connections**, 2021
- Multiple invited talks for **AI Illuminate**, **DPhi**, etc. on topics like **Zero Shot Text Classification**, **Practical Debugging for Data Science**