

Puneetha Pai B P

ML Engineer/Data Scientist

Currently a ML Engineer in ThoughtWorks. I have been in the industry for 3+ years. Joined TW as a developer and transitioned into data centric roles. Thus I have experience as an application developer, data science and Machine Learning engineer. I have good knowledge in python, statistical modelling, ML pipelines and automation.

SKILLS

- Languages: Python, Java, Scala, Shell script, SQL, js
- Data Science Practices: ML workflow automation, Model testing, Feature Engineering, Model Selection
- Dev Practices: TDD, CI/CD, Pair programming, Unit testing, Functional & automated testing
- Misc: noSQL, Performance Engineering, System Design
- Devops: Docker, Kubernetes, Serverless, AWS, terraform
- OS: Linux, MacOS

EXPERIENCE

ThoughtWorks, AI-Studio

Feb 2019 - PRESENT

Technologies Used: Python, NLP, NER, SpaCy, Prodigy, DVC, AWS, Athena, Terraform

Skill Ontology:

- Identify Skill to Skill relation using Market Basket Analysis over resume and CV data.
- Build a Skill-People graph to identify people with related skills for staffing needs.
- Skill and learning recommendations for people.

Superlative:

- Approach 1: Build a text classifier using Cosine Similarity to identify Superlative Phrases.
- Approach 2: Use SpaCy POS tagger to identify superlative adjectives. Then customise Spacy dependency parser to classify business centric superlative phrases.

NER

- Deploy a labeling solution, Prodigy, with client domain data for creation of labelled data set.
- Train a Named Entity Recognition model to identify entities like PERSON, GPE, PRODUCT and ORGANISATION using labeled data from Prodigy, which improved model recall from 64% to ~86%.
- Automate complete workflow of pre-process, train and model validation using DVC pipelines.

ThoughtWorks, Tesco Assessment

Dec 2018 - Jan 2019

Skills: Consulting, Pair Programming, Mentorship, Influencing, Agile

- Assess two retail solutions developed by the Tesco IT team, on code quality, architecture and identify reason for intensive effort on building new feature and enhancements
- Major anti-patterns discovered were
 - Separate teams for backend, front end, testing, devops and deployment with different cycles of integration and delay in process.
 - Over engineered backend with Cassandra and Elasticsearch, leading to parallel domain entities and mapping code. Any change to an entity causing a huge ripple effect.
 - Non reusable, custom/specific React UI components leading to duplicate effort.

- Other things like Dev Prod parity, build, CI-CD, test strategy, debuggability and branching.
- Finally onboarded the team in implementing the Software development best practices like
 - Practice Improvements: Strong Devops practice, anything as code analogy, automation of build, test, code quality adherence and Trunk based development
 - Process Improvements: Story Life Cycle management, Effective Meetings, task rotation

ThoughtWorks, University Trainer

Apr 2018 – Nov 2018

Skills: Teaching, Coaching, Mentoring, Counselling

- TW-University is a grad onboarding/training program.
- It aims to teach Software Development best practices by simulating a client project and sessions.
- I lead a team of 15 trainees, closely mentored 4 of them.
- Took sessions on practices like Agile Software Delivery, Consulting.
- Gave technical session on test strategy, blue green deployment, feature toggles, etc.

ThoughtWorks, Internal

Oct 2018 – Mar 2019

Technologies Used: Java, SpringBoot, Serverless (AWS Lambda), Postgres, CFN

- Built an online code evaluation platform for ThoughtWorks recruitment coding rounds
- Highly scalable, low cost serverless implementation, using AWS Lambda to evaluate the code submitted by the candidate
- Support for Java, Python, Golang and Javascript runtimes

ThoughtWorks, Data India Practice

Dec 2016 – Sep 2017

Technologies Used: Python, Keras, CNN, Genetic Algorithm, api.ai(chat bot)

- Trained a CNN deep learning model for object recognition to tackle inventory management problem
- Built a custom video processor to remove noise(eg: customers, staff) from inventory/store to focus on shelves, thus making it easier for prediction by the model.
- Built a chatbot for Trainline for getting schedule and booking tickets. Integrated it with Facebook messenger interface
- Implemented custom version of genetic algorithm for an optimisation problem

EDUCATION

WILP, BITS Pilani, Rajasthan

Dec 2018 – Present

M.Tech in Software Systems

Specialisation in DS and ML

CGPA: 9.7

SJCE, Mysore, Karnataka

June 2012 – May 2016

B.Tech in Electronics and Communications

Specialisation in DS and ML

CGPA: 9.37