

## Pankesh Bamotra

Mob : (412) 636 1222

Mail : pankesh.bamotra@gmail.com

## Education

### MCDS

**CARNEGIE MELLON UNIVERSITY, SCHOOL OF COMPUTER SCIENCE.**

MASTER OF COMPUTATIONAL DATA SCIENCE - DEC. 2016 - GPA **3.75**.

▷ *Machine Learning for Large Datasets.* ▷ *Search Engines.* ▷ *Natural Language Processing.* ▷ *Probabilistic Graphical Models.* ▷ *Multimedia Databases and Data Mining.*  
▷ *Deep learning.*

### B.Tech CSE

**VELLORE INSTITUTE OF TECHNOLOGY.** BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING - MAY 2013 - GPA **9.24**.

▷ *Data structures and algorithms.* ▷ *Operating systems.* ▷ *Linear algebra.* ▷ *Computer programming and problem solving.* ▷ *Digital Logic.* ▷ *Algorithm design and analysis.*

## Experience

### Coupage

**SR. DATA SCIENTIST - COMPUTER VISION.** APR '20 - PRESENT.

Working with visual intelligence team to develop models to identify counterfeit product listings and extract structured information from images

▷ *Working in partnership with catalog quality team to identify products that infringe trademarks and logos using object detection techniques.* ▷ *Working on an end-to-end model that extracts tabular information in images to enrich product catalog.*

### Coupage

**DATA SCIENTIST - COMPUTER VISION.** SEPT '18 - APR '20.

Worked with visual intelligence team to deliver a multi-category visual search engine for smoother customer shopping experience

▷ *Developed a metric learning based model to learn product embeddings for whole Coupage product catalog (100MM+ scale). Coupage price matching engine is built on this work.* ▷ *Built a sponsored search ads engine using semantic similarity model and FAISS. Served 100k+ ads per day.*

### Coupage

**RESEARCH ENGINEER - NLP.** FEB '17 - AUG '18.

Worked with Retail Systems team to ramp up product selection by optimizing operational efficiency through machine learning

▷ *Delivered a scalable product classification tool to organize products in hierarchical fashion. Improved product onboarding efficiency by 20%.*

### Autodesk

**SOFTWARE INTERN - BIG DATA ANALYTICS.** MAY '16 - AUG '16.

Worked with cloud platform team to migrate data from legacy system and ingest it into new Spark based platform

▷ *Worked with AWS, Spark, Oozie, Hadoop.* ▷ *New platform helped the company tune data ingestion, transport, and compute layers.*

### PayPal

**SOFTWARE ENGINEER.** JUN '13 - JUL '15.

Worked with Open Analytics Platform group to develop self-service database provisioning tool for analysts and developers

▷ *Refactored legacy code into a maintainable and testable stack.* ▷ *Developed plug and play solution for Qlikview and Teradata.* ▷ *Delivered SQL based solution to analyse clickstream data.* ▷ *Worked with one of the first agile teams at PayPal.*

### PayPal

**SOFTWARE DEVELOPMENT INTERN.** JAN '13 - JUN '13.

Worked with seller risk management to create early warning dashboards

▷ *Developed a SQL based solution to analyse buyer account habits.* ▷ *Early warning indicator reports used by teams across NA and EMEA for fraud analysis.* ▷ *Modeled post transaction risk mitigation.*

## Technical

|              |  |
|--------------|--|
| Advanced     | Python, Pandas, Scikit-learn             |
| Intermediate | PyTorch, PySpark, SQL, Java, Docker, C++ |
| Twiddling    | Kotlin                                   |

## Academic projects

|                        |  |
|------------------------|--|
| Info Retrieval         | <b>IMPLEMENTED A SEARCH ENGINE ON LUCENE.</b><br>▷ Supports exact match (Ranked Boolean). ▷ Best match using Okapi BM25 and Language model. ▷ Sequential dependency model. ▷ Query expansion. ▷ LeToR enabled.   |
| ML                     | <b>BRAIN fMRI IMAGE DATA ANALYSIS.</b><br>▷ Classification of fMRI images using SVM. ▷ Prediction of fMRI image voxels using ElasticNet Ridge Regression.  |
| NLP                    | <b>QUESTION GENERATION AND ANSWERING SYSTEM - BEST PROJECT AWARD.</b><br>▷ Wikipedia dataset. ▷ CLI interface to generate and answer questions. ▷ Generate semantically, grammatically, fluent questions and answers.                                  |
| Prob. Graphical models | <b>CONTEXT GENERATIVE RECURRENT NEURAL NETWORK LANGUAGE MODEL.</b><br>▷ Penn Treebank dataset. ▷ Optimized by Gensim, faster-rnnlm. ▷ Deep learning.   |
| Data mining            | <b>DATA MINING ON LARGE SCALE GRAPHS.</b><br>▷ Social graph datasets - Epinions, Gnutella. ▷ SQL based algorithms - Pagerank, K-core decomposition, SVD, triangle counting. ▷ Pattern analysis on static and temporal graphs using power law analysis. |
| Capstone-1             | <b>FINANCIAL PRICE PREDICTION FROM EARNING CALL TRANSCRIPTS.</b><br>▷ Predicts directional movement of stock on next trading day. ▷ Predicts change in stock price on next trading day. ▷ Analysis purely using NLP and historical stock prices.       |
| Deep learning          | <b>BEAUTIFYING HANDWRITING USING NEURAL NETWORKS..</b><br>▷ Applies extracted style to handwriting from existing font. ▷ Extraction of style using VGG19. ▷ Maintains text legibility and aesthetics.  |
| Capstone-2             | <b>STANFORD QUESTION ANSWERING DATASET CHALLENGE..</b><br>▷ Reading comprehension task. ▷ 100k+ question-answer pairs. ▷ Used fusion model combining NLP and deep learning techniques.   |

## Publications

|                           |   |
|---------------------------|---|
| IOSR Journal<br>(Sep '12) | <b>SECURING GROUP COMMUNICATION IN PARTIALLY DISTRIBUTED SYSTEMS.</b> SYMMETRIC KEY CRYPTOGRAPHY, KEY DISTRIBUTION. |
| IJSCE<br>(Nov '12)        | <b>SECURE TRANSMISSION OF GRAYSCALE IMAGES USING DISCRETE FOURIER TRANSFORM.</b> IMAGE PROCESSING, SECURITY.        |
| IJARCSSE<br>(Dec '12)     | <b>IMAGE ENCRYPTION USING PIXEL SHUFFLING.</b> ENCRYPTION.  |
| IJAIS, FCS<br>(Jan '13)   | <b>LRUD SHIFT BASED ENCRYPTION METHOD USING MATRICES.</b> ENCRYPTION, SECURITY, LINEAR ALGEBRA.                     |
| IJAIS, FCS<br>(May '13)   | <b>HADOOP AND RISK ANALYTICS.</b> ANALYTICS, FINANCE, BIG DATA.   |