Hossein Azizi, PhD

Montreal, QC, Canada | (514) 5828756 | hosseinphy@gmail.com| Github| linkedin

Technical Skills

Stats & Machine Learning:

- **Descriptive Statists:** Bayesian Statistics | Time-series Analysis | A/B Testing | Hypothesis Testing | KS Test | ANOVA
- **Supervised Learning:** Linear & Non-linear Regression | Random Forest | SVM | Neural Networks (CNN, RNN, LSTM)
- **Unsupervised Learning:** Clustering (K-Means, GMM, Hierarchal) | PCA
- **Reinforced Learning:** UCB | Thompson Sampling

Coding:

- **Languages:** Python | C/C++ | Spark | MATLAB | CUDA | HTML | Flask | shell scripting
- **Libraries:** Pandas | NumPy | SciPy | Scikit-Learn | Keras | TensorFlow | Nltk | spaCy | Beautifulsoup | Networkx | Matplotlib | Bokeh | Altair | Seaborn | ParaView
- **Algorithm Design:** OpenMp | MPI | multi-Threads | Monte Carlo Simulation | Image Analysis | Fourier Analysis | Dynamic Adaptive Mesh | Algorithm Optimization | PDEs
- **Continuous Integration/Analysis pipeline:** Git | make, Bash
- **TDD:** Unit tests | Integration tests | Acceptance tests

Data Management: SQL (Postgres, SQLite) | AWS (Lambda, S3, Redshift, EMR)

Vertical Knowledge: Recommendation Systems | Market Segmentation | Digital Strategy for Materials Design | Functionality Optimization | Energy | 3D Metal Printing | Combustion | Process Optimization & Automation | Computational Physics | Pattern Recognition

Experience

1. **Challenge:** To build a model to predict the tag of questions in a Q&A website.
1. **Solution:** Parsed, cleaned and processed a 10 GB set of XML files. Analyzed long-term user behavior using spark SQL. Implemented a spark ML pipeline to train a word2vec model and classification model to predict the tag of a question from its text.
2. **Challenge:** To predict the rating of Yelp review data based on the text of reviews.
2. **Solution:** Built & train a bag_of_words model (TfidfVectorizer + Ridge) on 250,000 reviews to predict star ratings. Features extracted by normalizing (tf-idf weighting) single words and bi-grams in the corpus with lemmatization and stop words supplied by spaCy. Ran sentiment analysis on the reviews by deploying naïve Bayes model.
3. **Challenge:** To assemble a social graph of the social connections of New York's social elite.
3. **Solution:** Crawled and cleaned ~ 96,000 photo captions from more than 1000 web pages. Parsed list of names using Nltk (POS tagging, Chunking/Chinking) and regex. Assembled the social graph using networkx and determined most influential/popular people within the group and their best friends.
4. **Challenge:** To create a movie box-office revenue prediction model.
4. **Solution:** EDA (data distributions/correlations); Feature engineering (created features based on text and categorical data); Built & trained a random forest classifier for multiclassification: used ensemble under sampling to combat unbalanced data, precision = 0.69, recall = 0.72.
5. **Challenge:** To build an optimized neural networks model to perform image classification.
5. **Solution:** Transferred 60,000 images, each belonging to one of ten classes. Used Keras to build & train a convolutional model with Dropout and Early Stopping to avoid overfitting, validation accuracy 70%. Improved the accuracy by 23% by using transfer learning from Inception model.

Hossein Azizi
☎ (514) 5828756 | ✉ hosseinphy@gmail.com | Github | Linkedin

6. **Challenge:** To design an application to forecast fitness product online sales (Capstone project).
6. **Solution:** Scraped historical data for 1000 unique products; Handled blocking and scraping from ajax websites; Feature engineering (creating new features based on inventory history, detecting & removing outliers based on percentile, bucketizing price & dates); Build a Decision-Tree based regression model to predict inventory.
7. **Challenge:** To design a numerical model to simulate & optimize 3D Metal Printing process of mechanical parts.
7. **Solution:**
    a. Modelled crystallization process, collaborated with an engineering team to develop realistic laser thermal profile.
    b. Designed & implemented a high-performance computing algorithm for large-scale simulations, resulting in 35x speedup across 1024 cores by applying MPI techniques.
8. **Challenge:** To train a ML algorithm to predict pattern formation in stochastic reaction-diffusion systems.
8. **Solution:** Executed 40+ simulations of the process using 2000+ cores; Initiated process automation that required to collect, interpret, and visualize more than ~10TB of experimental and simulation data to develop training dataset. Accelerated processing time 7-fold (from 14 to 2 days).
9. **Challenge:** To model pattern formation in randomly structured materials.
9. **Solution:** Developed parallel 2D Monte Carlo code in C to generate variety of randomly structured materials. Improved accuracy of material's characterization by 20%. Created simulation of heat-mass transfer process to assist in design new porous materials, with potential use in energy-efficient products.

Work Experience

Marine Additive Manufacturing Centre, University of New Brunswick, **Feb 2018 - Aug 2021**
Postdoctoral Research Fellow

- Supervised and trained a team to apply simulation to improve decision making process and parts' quality (Exper-7).
- Initiated a project to use pattern formation predictions to assist in design of new Functionally Graded Materials (Exper-8).
- Worked with a team to customize and optimize pre-existing parallel 2D & 3D open-source simulation code, resulting in 100-fold acceleration and reduction of computational cost and memory usage. The software currently being used by research labs in Canada, US, Finland, and China.

Department of Mechanical Engineering, McGill University **Sep 2017 - Jan 2018**
Research Associate

- Supervise & train a student to analyze and interpret research data (Exper-9)

Education

Data Scientist Certification, The Data Incubator Fellowship Program | Aug 2021

PhD in Computational Physics | McGill University | Sep 2017

Data Science Workshops & Courses

Udemy:

Python Programming Masterclass (2021); Machine Learning A-Z: Hands-On Python & R in Data Science (2021)

McGill High Performance Computer Centre, Montreal, Canada:

Advanced & parallel Python (2016); Introduction to CUDA/GPU (2016); Data Intensive Computing (2016); Advanced MPI to create parallel software (2014); Scientific Visualization, ParaView (2014)

Publications

Authored 4 papers, presented in 3 prestigious conferences, invited speaker in 2 symposiums. Google Scholar