

---

# Hossein Azizi, PhD

---

Montreal, QC, Canada | (514) 5828756 | [hosseinphy@gmail.com](mailto: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 | Boosting | Neural Networks (CNN, RNN, LSTM)
- **Unsupervised Learning:** Clustering (K-Means, GMM, Hierarchal) | PCA
- **Reinforced/Active Learning:** UCB | Thompson Sampling | Gaussian process

### Coding:

- **Languages:** Python | C/C++ | Spark | MATLAB | CUDA | HTML | Shell Scripting
- **Libraries/ Frameworks:** Statsmodels | Pandas | NumPy | SciPy | Scikit-Learn | Keras | TensorFlow | Nltk | spaCy | BeautifulSoup | Networkx | Matplotlib | Bokeh | Altair | Seaborn | ParaView | Flask | Django
- **Algorithm Design:** OpenMp | MPI | multi-Threads | Monte Carlo Simulation | Image Analysis | Fourier Analysis | Dynamic Adaptive Mesh | Algorithm Optimization | PDEs
- **Tools:** Git/Bitbucket | Docker | Make | Bash
- **TDD:** Unit tests | Integration tests | Acceptance tests

**Data Management:** SQL (Postgres, SQLite, SQLAlchemy) | AWS (S3, Redshift, EMR, EC2)

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

## Work Experience

**CanmetMATERIALS, Natural Resources Canada,**

**Jan 2022 – Present**

### Programmer/ Data Scientist (Full time)

- Designed & tested archive databases; Implemented & tested backup strategies for disaster recovery (Exper-1)
- Developed archiving solution for autonomous optimization system:
  - Implemented & tested an archive executor application using Django.
  - Integrated archive executor into data orchestration system, run asynchronous tasks with celery and RabbitMQ.
- Created a backend API for front-end tools, using Django REST framework.
- Development of autonomous multi-parameter optimization application:
  - Collaborated with software development team to debug, test, and execute software codes.
  - Deployed & documented a walkthrough tutorial for the data orchestration framework.
  - Provided support to experiment team to design optimization workflow for different experiments.
  - Designed & implemented machine learning algorithms and simulations to inform decision making processes.
- Conducted numerical phase-field simulations to model hybrid additive manufacturing of dilute ternary Al-based alloys

**Marine Additive Manufacturing Centre, University of New Brunswick,  
Postdoctoral Research Fellow**

**Feb 2018 - Aug 2021**

# Hossein Azizi

☎ (514) 5828756 | ✉ [hosseinphy@gmail.com](mailto:hosseinphy@gmail.com) | [Github](#) | [Linkedin](#)

- Supervised and trained a team to apply simulation to improve decision making process and parts' quality (Exper-11).
- Initiated a project to use pattern formation predictions to assist in design of new Functionally Graded Materials (Exper-12).
- 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**  
**Research Associate**

**Sep 2017 - Jan 2018**

- Supervise & train a student to analyze and interpret research data (Exper-13).

## Experience

1. **Challenge:** To design & test databases for a Data Orchestration Service.
1. **Solution:** Designed an efficient archive database according to system requirements using PostgreSQL. Conducted performance tests on the database. Recommended modifications in the existing databases to improve the performance of the systems. Provided documentation on database structures, specifications and functionalities.
2. **Challenge:** To build scalable pipelines that can ingest, update and process job posting data using existing AWS stack.
2. **Solution:** Produced a bash script to transfer entire image of the original dataset (30GB) and following weekly updates (> 1GB) to the EC2 instance over SFTP. Transferred data in parallel from the EC2 instance to the Redshift cluster via SSH protocol (using COPY command and ssh-manifest file in s3). Wrote SQL queries and functions in the Redshift query editor to update and process job posting data.
3. **Challenge:** To build a prototype recommendation system for a video sharing site.
3. **Solution:** Downloaded & parsed user/video data from Duban datasets. Designed & deployed a session-based KNN algorithm to recommend next videos to a user. The algorithm was optimized using KNN-elements sampling method.
4. **Challenge:** To design a model for novelty detection of bike ridership time series data.
4. **Solution:** Built and trained a full background model to predict seasonality and drift, using Scikit-learn transformers & pipelines,  $R^2 = 0.7716$ . Added windowed features to account for the short-term trends using the exponentially-weighted moving average (EWMA), improve the accuracy of the full background model by ~5%. Detected and interpreted anomalous observations by analyzing the z-score of the residuals (actual minus predicted values).
5. **Challenge:** To predict the rating of Yelp review data based on the text of reviews.
5. **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.
6. **Challenge:** To build a model to predict the tag of questions in a Q&A website.
6. **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.
7. **Challenge:** To assemble a social graph of the social connections of New York's social elite.
7. **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.
8. **Challenge:** To create a movie box-office revenue prediction model.
8. **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.
9. **Challenge:** To build an optimized neural networks model to perform image classification.

## Hossein Azizi

☎ (514) 5828756 | ✉ [hosseinphy@gmail.com](mailto:hosseinphy@gmail.com) | [Github](#) | [Linkedin](#)

9. **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.
10. **Challenge:** To design an application to forecast fitness product online sales ([Capstone project](#)).
10. **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.
11. **Challenge:** To design a numerical model to simulate & optimize 3D Metal Printing process of mechanical parts.
11. **Solution:**
- Modelled crystallization process, collaborated with an engineering team to develop realistic laser thermal profile.
  - Designed & implemented a high-performance computing algorithm for large-scale simulations, resulting in 35x speedup across 1024 cores by applying MPI techniques.
12. **Challenge:** To train a ML algorithm to predict pattern formation in stochastic reaction-diffusion systems.
12. **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).
13. **Challenge:** To model pattern formation in randomly structured materials.
13. **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.

## 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](#)