# Reza Soleymanifar
## Data Scientist

+1 217 819 9898 | reza@soleymanifar.com | linkedin.com/in/soleymanifar | github.com/rezasoleymanifar | soleymanifar.com

## EDUCATION

**University of Illinois at Urbana-Champaign** — Urbana, IL
*Ph.D. in Industrial Eng. specialization in data analytics, thesis on temporal clustering* — *Aug. 2018 – July 2022*

**Sharif University of Technology** — Tehran, Iran
*M.Sc. in Systems Eng. thesis on reinforcement learning for revenue management* — *Aug. 2014 – May 2016*

**University of Tehran** — Tehran, Iran
*B.Sc. Industrial Eng.* — *Aug. 2009 – May 2013*

**Courses**: Deep learning, machine learning, data analytics, algorithms, nonlinear optimization.

## SKILLS

PyTorch, Keras, TensorFlow, Python, Bash, Git, SQL, AWS SageMaker, OpenCV, BigDL, scikit-learn, Numpy, pandas, OOP

## EXPERIENCE

**Computer Vision Scientist Intern** — May. 2021 – Aug. 2021
*Cargill* — *Champaign, IL*

- Created an end-to-end computer vision solution for detecting workers in hazard zones and raising an alarm if needed. Project was acclaimed as a *foundational* contribution to data science team and will be deployed in Cargill facilities in US.
- Used parallel programming and pretrained YOLOv3 to simultaneously download training data from plant IP cameras (final dataset: +3000 annotated images).
- Used transfer learning for 5 state of art object detection models pre-trained on COCO 2017 dataset. Trained and deployed models on AWS SageMaker. Used Docker to create prebuilt GPU training containers on AWS ECR.
- Used mAP (IoU $\geq$50%), frame rate and inference time to benchmark models and picked EfficientDet-D1 as the winner.
- Used OpenCV KCF tracker to track the motion trajectory of workers in the plant.

**Data Scientist Intern** — May 2020- Aug. 2020
*Epsilon* — *Chicago, IL*

- End-to-end retail product *image classification* project for marketing. Will be used for personalizing online ads to increase click through rates.
- Used Epsilon's marketing database: +2,000,000 images, +200 product categories from Lowe's, Cabelas, Walgreens, and other major retailers.
- Created the training dataset using *PySpark* pipeline and RegEx to extract annotations for raw images. Used Hadoop HDFS commands for file operations.
- Trained *CNN* model over *Hadoop* clusters using BigDL package (distributed deep learning). Used image augmentation, spatial drop-out, Gaussian noise and early stopping to regularize the model.
- Fine-tuned kernel sizes, strides, and layers using k-fold cross validation. Achieved significant top-1 classification accuracy improvement (+50%) over the baseline model.

## PROJECTS

- **Atari AI** | *Keras-rl, gym*: Compared actor-critic, policy-gradient and deep q-learning algorithms for training an agent to play the MsPacman-v0 game in OpenAI Gym. Used RTX 2060 GPU to train the agent.
- **Seq2Seq** | *PyTorch, TorchText, spaCy*: Developed a sequence to sequence encoder-decoder LSTM model for German to English translation. Used Multi30K dataset to train the model.
- **GAN** | *TensorFlow, TensorBoard*: Trained a generative adversarial network to perform image super-resolution on low resolution images. Used Yahoo MirFlickr25k dataset to train the model.
- **Speech technology** | *librosa, scipy.io, Keras*: Designed a ConvNet that understands word utterances. Used tensorflow Speech Commands Dataset (65,000+ utterances, 30 words) to train the model, 71% test accuracy.
- **Kaggle Competition** | *sklearn, pdpipe, pandas*: Designed stacking ensemble of random forest, extremely randomized/gradient boosted trees, and a multilayer perceptron to predict the prices of homes in the Ames Housing dataset. Achieved RMSLE of 0.0061
- **Software Development**: Contributed 20k+ lines of VBA code to MTN in-house budgeting analytics software, CapEx TS that is operational to present day with 20 concurrent users.

## PUBLICATIONS

[1] SOLEYMANIFAR, R. Lp based approximate dynamic programming model to address airline overbooking under cancellation, refund, and no-show. *Yugoslav Journal of Operations Research 29 (1)* (2018), 113–133.

[2] SOLEYMANIFAR, R., AND BECK, C. Rcp: A temporal clustering algorithm for real-time controller placement in mobile sdn's. *pre-print* (2021).

[3] SOLEYMANIFAR, R., AND KHAYYER, A. A purely data-driven approach to airline overbooking under realistic constraints using deep q-learning. *pre-print* (2021).

[4] SOLEYMANIFAR, R., SRIVASTAVA, A., BECK, C., AND SALAPAKA, S. A clustering approach to edge controller placement in software defined networks with cost balancing. *21st International Federation of Automatic Control World Congress* (2020).