

HUSEIN ZOLKEPLI

Email: husein.zol05@gmail.com

Gitlab: <https://gitlab.com/huseinzol05>

Website: <https://huseinhouse.com>

Github: <https://github.com/huseinzol05>

Kaggle: <https://www.kaggle.com/huseinzol05>

Linkedin: <https://www.linkedin.com/in/husein-zolkepli/>

A software engineer focused on machine learning, devops, big data and data science. His hobby is chilling watching his flowers at his balcony.

Work Experience

Chief Data Scientist

Bitcurate CO, Kuala Lumpur, MY

February 2019 - Now

Crypto market intelligence. Helping investors with real-time insights through machine learning and predictive analytics.

- Provide reliable durability, flexible publish-subscribe/queue for various exchanges and pairs cryptocurrencies that scales well using Apache Kafka to increase more power to distributed real time data processing.
- Increase 300% distributed real time data processing by using Apache Storm, dynamic consumers for Apache Kafka to provide almost real time cryptocurrencies-sentiment correlation.
- Achieve better predictive modelling to forecast future trends and probabilistic confident level for forecasted trends using Linear Regression, ARIMA and LSTM models on Apache Airflow.
- Reduce 95% deployment time by fully automated continuous integration deployment using Jenkins.

Chief Technical Officer

Mesolitica Sdn Bhd, Kuala Lumpur, MY

December 2018 - Now

Social media analysis and open source startup to support development of Malaya, Natural-Language-Toolkit for Bahasa Malaysia, and building massive Bahasa Malaysia corpus.

- Achieve better topics semantic definitions and topics similarities daily updated by using Apache Airflow for scheduled crawlers to crawl local news and Telegram.
- Crawl blogspot, Instagram, property websites to prepare wider platform data points for data scientists.
- Achieve very accurate text mining results by deploying deep learning models build on top of Malaya, <https://malaya.readthedocs.io/> using Gunicorn and Docker Swarm.
- Provide reliable durability, flexible publish-subscribe/queue for various social media data points that scales well using Apache Kafka to increase more power to distributed real time data processing.
- Increase 500% distributed real time data processing by using Apache Storm, dynamic consumers for Apache Kafka.
- fully engaged in day-to-day operations with internships on how to manage Docker Swarm and Elasticsearch clusters.
- Reduce 95% deployment time by fully automated continuous integration deployment using Jenkins.

Data Engineer

Omnytics Co, Shoppr Sdn Bhd, Kuala Lumpur, MY

August 2018 - November 2018

The leading fashion & beauty market intelligence platform that will grow your business with tools providing actionable insights for sales, marketing and merchandising.

- Refactor Elasticsearch clusters and indices to make it more efficient and 75% less space used.
- Serve aggregation APIs for alternate sales index, sales forecasting, trend forecasting, market growth and market insight to help our clients make better decision.
- Improve big data pipelining and scheduling using Luigi Spotify.
- Cover 95% of unit tests to improve accuracy and flow of the APIs and system provided.

Data Scientist

XFERO Sdn Bhd, MaGIC, Cyberjaya, MY

February 2018 – October 2018

- Reduce Implementation failure, reflections and recovery time by changing all microservices into Docker Container.
- Improve coordination to unify development code and continues integration from various departments using CircleCI.
- Help agriculture companies do better decision on yield using real time computer vision probability inference on fruits quality.
- Help drone community by created Tensorflow deep learning interface to drone DJI Tello.

Executive Robotics (AI)

Robopreneur Sdn Bhd, Futurise, Cyberjaya, MY

March 2018 – August 2018

brings the future by empowering capabilities of service robotics for use in human's environment, solve human problems and become human's partner.

- Develop text to speech engine for small embedded devices, using MQTT and Websocket to improve Robopreneur Humanoid chat system and smarter robot.
- Develop face recognition, objection recognition, image semantic engines for Robopreneur IOT camera devices to reduce error on road image understanding and recognize environment objects for self-driving car.
- Develop advanced face analysis (roll, pitch, yaw, emotion, gender, recognition, age, eyes open probability, mouth open probability) in a single inference time to improve Robopreneur Humanoid human interaction.

Machine Learning Engineer

Digital Hill Sdn Bhd, Kuala Lumpur, MY

August 2017 – July 2018

specialize in solving multitude issues in the large and small perspectives scale of economics, business and strategic communications using data analytics.

- Improve local social media understanding for Bahasa Malaysia and English by develop bilanguage sentiment analysis machine learning model.

- Improve political campaign by study social network and political landscape for DUN and parliaments during Malaysia General Election 14.
- Crawl blogspot, Instagram, twitter and local news to understand millennial political landscape and build monte carlo election simulation to give different perspectives to Wakil Rakyat.

Open-source Developer

Devcon Community, MaGIC, Cyberjaya, MY
February 2017 – Jan 2019

A small community focused on open source projects and community services for digital transformation.
<https://github.com/devconx>

- Conduct free classes related to deep learning, data science, data engineer, modern DevOps to help local students prepare to different software industries.
- Maintain Malaya Repository, Bahasa Malaysia NLP library with deep learning Tensorflow to bring Malaysia towards to Industry 4.0.

Founder

Soaning, Vega Residensi 1, Cyberjaya, MY
May 2017 – Feb 2018

- Develop stock engine that able to crawl selected hyper-parameters from Internet to predict specific stock using LSTM model to help buyers make better decision on stock trading.
- Develop and serve API that able to understand local sentiment analysis to help local social media analytics make better decision on sentiment understanding.
- Develop screening API for Github, able to classify developer personalities and technology concern to help companies make better decision to hire developers based on his github repositories.
- Develop screening API for CV / Resume, able to classify MBTI and developer personalities.
- Develop intelligent vision agriculture system able to study attractive attributes of flower and healthy percentage from crop to reduce human error and cost.

Full Stack

Swift Enterprise, Kedah, MY
Sept 2016 – Feb 2017

- Develop Facebook Ads Management System using PHP, MySQL, JQuery, Bootstrap and ChartJS to automate targeting ads on certain location and landscape for Facebook pages.

Open-source Projects

Malaya-Dataset

Text corpus for Bahasa Malaysia.

<https://github.com/huseinzol05/Malaya-Dataset>

Open sourced categorized articles, audience nationality, dependency parsing, 25k unique words dictionary, emotion, entities, gender, irony, karangan sekolah, language detection, categorized news, Twitter-based sentiment, news-based sentiment, multi domain sentiment, part of speech, polarity, political landscape, sarcastic based on news headline, stemmer, subjectivity, toxicity and subtitle.

Malaya

Natural Language Toolkit for Bahasa Malaysia, powered using deep learning models.

<https://github.com/huseinzol05/Malaya>

Supports news crawler, Entities Recognition, Language Detection, Normalizer, Num2Word, POS Recognition, Sentiment Analysis, Spelling Correction, Stemmer, Summarization, Topic Modelling, Topics and Influencers Analysis, Toxicity Analysis and Semantic Analysis.

Tello-Python

Tello packets prototype integrated with Tensorflow object and face detection models.

<https://github.com/DevconX/Tello-Python>

Reversed engineered Tello UDP streaming application to Python friendly and integrated with Tensorflow object and face detection during flight.

Deep-Learning-Tensorflow

Gathers Tensorflow deep learning projects.

<https://github.com/huseinzol05/Deep-Learning-Tensorflow>

Implementing various Convolutional Neural Network, Recurrent Neural Network, Feed-forward Neural Network, Sequence-to-Sequence, Hybrid, Bayesian optimization model, transfer learning, tensorboard, tf-distributed, tf-serving, tensorboard debugger, multiprocessing tfrecords, Regression and Generative Adversarial Network.

NLP-Models-Tensorflow

Gathers machine learning and tensorflow deep learning models for NLP problems.

<https://github.com/huseinzol05/Deep-Learning-Tensorflow>

Implementing text classification, chatbot, neural machine translation, embedded, POS-tagging, Entity-tagging, Question-Answers, Text-matching, Summarization, Attention model, Stemming, OCR, Language Detection, and Miscellaneous. Contain more than 220 deep learning models.

Stock-Prediction-Models

Gathers machine learning and deep learning models for Stock forecasting, included trading bots.

<https://github.com/huseinzol05/Stock-Prediction-Models>

Implementing stacking (machine and deep learning models) models, deep-learning models, trading agents, stock data analysis and fashion trending with cross-correlation. Covered LSTM, GRU, Sequence-to-Sequence, Attention model, and Differentiable Neural Computer.

Tensorflow-JS-Projects

Web projects using Tensorflow JS, Plotly, D3, Echarts, NumJS, and NumericJS

<https://github.com/huseinzol05/Tensorflow-JS-Projects>

Implementing tensorboard-like histogram on MNIST dataset, decision boundaries for IRIS using PCA and SVD, Linear, Polynomial, Ridge, Lasso and Elasticnet Regression, Stock forecasting and investment simulation with distribution study, Malaysia Export products forecasting and trading agent using Evolution Strategy.

Reinforcement-Learning-Agents

Gathers machine learning and deep learning models for Reinforcement Learning.

<https://github.com/huseinzol05/Reinforcement-Learning-Agents>

Implemented reward based for Evolution Strategy, policy gradient, q-learning, double q-learning, recurrent q-learning, double recurrent q-learning, dueling q-learning, dueling recurrent q-learning, double dueling recurrent q-learning, actor-critic, actor-critic dueling, actor-critic recurrent, actor-critic dueling recurrent and async q-learning.

Self-Driving-Car-Engines

Gathers signal processing, computer vision, machine learning and deep learning for self-driving car engines.

<https://github.com/huseinzol05/Self-Driving-Car-Engines>

Implemented Signal processing (1D smoothing, 2D smoothing, convolution 2 signals, pass-filters), Simple straight lane detection, Steering suggestion, Multi-lane detection, Multi-lane angle, Distance + Angle for object detection, Traffic light detection, Road Segmentation, Plate detection, Dynamic count lane detection and Gradient Smoothing.

Python-DevOps

Gathers Python stack for DevOps, everything is Docker!

<https://github.com/huseinzol05/Python-DevOps>

AutoPEP8, Pytest Flask, Flask, MongoDB, Flask Rest API, Postgres, Elastic Search, Luigi Spotify, Kibana, Distributed SocketIO with Redis, ELK Flask, Flask Hadoop, Mlfow Nginx, Flask Kafka, Flask Hive Hadoop, Pyspark Jupyter Hadoop.

Gather-Tensorflow-Serving

Gather how to deploy tensorflow models as much I can.

<https://github.com/huseinzol05/Gather-Tensorflow-Serving>

Implemented Object Detection using Flask SocketIO for WebRTC, Object Detection using Flask SocketIO for opencv, Speech streaming using Flask SocketIO, Multiple Inception with Flask using EC2 Docker Swarm + Nginx load balancer, Text classification using Hadoop streaming MapReduce, Text classification using Kafka and Text classification on Distributed TF using Flask + Gunicorn + Eventlet.

Machine-Learning-Data-Science-Reuse

Gathers machine learning and data science techniques for problem solving.

<https://github.com/huseinzol05/Machine-Learning-Data-Science-Reuse>

Implemented data preprocessing, natural language processing, suggestion engine, image processing, signal processing, stacking model, stochastic study, data visualization, markov, english-text normalization and sound processing.

Machine-Learning-Numpy

Code Machine learning models without any frameworks, Numpy only.

<https://github.com/huseinzol05/Machine-Learning-Numpy>

Implemented feed-forward, vanilla recurrent, lstm recurrent, gru recurrent, convolutional, batch-normalization, dropout, regularization, neuro-evolution, evolution strategy, clustering, decomposition, probabilistic, regression, trees based, timeseries and signal processing.

Neural-Network-Multilanguages

implement Artificial Neural Network on different languages.

<https://github.com/huseinzol05/Neural-Network-Multilanguages>

implement Gradient Descent Feed-forward and Recurrent Neural Network on different languages, only use vector / linear algebra library. Supported Julia, Python, Javascript, Go, C++, Ruby and PHP.

Pyspark-ML

Gathers data science and machine learning problem solving using PySpark and Hadoop.

<https://github.com/huseinzol05/Pyspark-ML>

Logistic regression, multinomial classification, topic modelling, word vector, principal component analysis, deep learning Tensorflow on Spark Cluster.

Hackathon

NASA Hackathon 2017	create a system to forecast earthquake.
GRAB Hackathon	create a system to predict traffic density.
Malaysia Airline Hackathon	create in-flight chat-bot system.
Advoc8 Hackathon	create a system to give multi-tag for pictures.
MYDD 2017	create a system to detect emotion and play music based on the emotion.
Selangor Smart City 2017	create smart vandalism detection system.
MAMPU Data Hackathon 2017	create correlation study system.
TADHACK 2017	create crop and flower system.
FACE 2017	create smart house system.
REKATHON 2017	create in-house lidar system.
FisHackathon 2018	create fishermen forecasting system.
Makerthon 2018	main judge.
AWS Hackathon 2018	create Text Analysis for large company commerce dashboard.
MYDD 2018	create IOT tracking for turtles.
ZooHackathon 2018	create realtime animal detection dashboard.
Genting Programming	

Academic Qualification

Sijil Pentaksiran Menengah MARA Junior Science College Taiping, Malaysia (2013)

Skill

Programming Language

C, C++, Java, Julia, Go, Ruby, Python, Javascript, PHP, C#

RDBMS

Postgres, MySQL, Oracle, RedShift, Hive

Apache Product

Hadoop, Kafka, Spark, Hive, Storm, Airflow

NoSQL

Redis, MongoDB, DynamoDB, Elasticsearch

Cloud Service

Amazon, Google Cloud

Game Engine

Unity 3D (C# API), Unreal Engine 4 (Blueprint and C++ API)

ETL / Piping

Luigi spotify, Apache Airflow

Python big data

Dask, luigi, pyspark, hadoop streaming

Python data science / machine learning stack

Tensorflow, Matplotlib, seaborn, Scikit-learn, networkx, plotly, opencv

DevOps

Docker, Git, Kubernetes, Docker swarm, Docker compose, Gitlab CI, Circle CI, Travis, Jenkins

Language Proficiency

English: Good in writing and speaking.

Malay : Fluent in writing and speaking.

Referee

Referee 1:

DR. NORSHUHANI ZAMIN

University Malaysia of Computer Science and Engineering,

Precint 1, Putrajaya, 62000,

Malaysia

019-6626772

norshuhani@unimy.edu.my

Referee 2:

DR. AHMAD FAZREEN BAHARUDEN

DIGITAL HILL SDN BHD (833721-H)

Suite 3.2, Level 7, Putra World Trade Centre

41 Jln Tun Ismail, 50480 Kuala Lumpur, Malaysia

03-40444208

fazreen@digitalhill.com.my

Referee 3:

FAIZUL HIDAYAT BIN MOHD ARDANI

XFERO SDN BHD (1168264-H)

MaGIC, Cyberjaya, Malaysia

013-5848137