

Zehni Khairullah

JUNIOR DATA SCIENTIST · FULL STACK DEVELOPER

Van Polanenpark 193, 2241 SK, Wassenaar, The Netherlands

☎ (+31) 06 23 57 54 12 | ✉ mzkhairullah@gmail.com | 🏠 zehni.github.io | 📱 zehni | 🌐 zehni1

Summary

Junior data scientist with 1+ years of experience in course work and internships along with 3+ years of experience in full stack development and product ownership through internships and freelance jobs. I bring in the business and technology knowledge to solve complex business problems by leveraging data. I am an entrepreneurial team player, risk taker and a self-motivated individual. I specialize in predictive analysis, machine learning, deep learning, full stack development and product development.

Relevant Experience

Irdeto B.V.

Hoofddorp, The Netherlands

DATA SCIENCE INTERN

Aug. 2018 - Dec. 2018

- Built and deployed overall services such as Kubernetes & Kubeflow utilizing Docker containers and several AWS stacks (including EC2, S3, IAM) focusing on high availability, fault tolerance and auto-scaling of deep learning models
- Collected and visualized metrics from Kubernetes cluster by setting up Istio metrics, Prometheus and Grafana, which helped monitoring the activity of the deployments
- Examined the ability of training a deep learning model on Kubernetes using Kubeflow, focusing on accelerated distributed training and the availability of an up to date model
- Assisted and contributed to visualizing site-blocking (time series data) in an understandable way for business personnel

Baaz, Inc.

Doha, Qatar

BUSINESS ANALYST INTERN

Jun. 2017 - Aug. 2017

- Captured business requirements in the form of epics, features and user stories making sure the acceptance criteria for the user stories are clear
- Acted as a liaison on day to day basis with the product owner, user experience specialist and development team.
- Collaborated with the testing team & insured requirements were implemented as proposed
- Prepared and reviewed monthly sprints with the scrum master & clarified all user stories to be implemented

Baaz, Inc.

Doha, Qatar

BUSINESS SOFTWARE ANALYST INTERN

Dec. 2015 - Feb. 2016

- Helped understand the system's features using user stories by acting as a liaison between technology and business
- Defined business rules, gathered system's requirements and new features to gain competitive advantage by contributing to the business and development team
- Enhanced the user experience for potential users by documenting the current features of the system

Skills

Programming	Python, R, Matlab, Java
Web	Django, HTML5, CSS3, JavaScript, jQuery
Frameworks & Databases	Tensorflow, Keras, Kubernetes, Kubeflow, PostgreSQL, MySQL, MonetDB, MongoDB
OS & Platforms	MacOS, Linux, AWS, MS Azure, Google Cloud, Grafana, Prometheus
Languages	Arabic, English, Dutch(B1)

Education

Leiden University

Leiden, The Netherlands

MSC IN COMPUTER SCIENCE (DATA SCIENCE SPECIALIZATION)

Sept. 2017 - Exp. Jul. 2019

- **Key Courses:** Advances in Data Mining, Multi-objective Optimization, Evolutionary Algorithms, Neural Networks, Linear Models & Generalized Linear Models, Information Retrieval & Text Analytics, Multivariate Analysis and Statistical Learning Theory

Carnegie Mellon University

Pittsburgh, PA, USA

BSC IN INFORMATION SYSTEMS (EXTENSIVE BUSINESS COUREWORK)

Sept. 2012 - May 2016

- University Honors (Major GPA 3.82/4.0)
- Senior Leadership Award
- **Key Courses:** Web Application Development, Business Technology Consulting, Regression Analysis, Decision Making Analysis, Organizational Behaviour, International Management, Cross Culture Business Communication and Digital Marketing

Selected Projects

Cardinality Estimation Algorithms Review and Comparisons

PROBLEM: COUNTING THE NUMBER OF UNIQUE ELEMENTS IN A BIT STREAM WHERE STORAGE IS NOT ENOUGH.

- Flajolet-Martin Algorithm also known as Probabilistic Counting with Stochastic Averaging (PCSA) and Flajolet-Durand Algorithm or LogLog algorithm were compared
- The Relative Approximate Error (RAE) was used as the evaluation metric
- 450+ experiments were conducted along with Monte Carlo simulations
- The project was implemented in Python

Real-world Recommendation Systems

PROBLEM: BUILD RECOMMENDATION SYSTEM FOR THE 1M MOVIELENS DATA SET

- Naive models were developed as global average of all ratings, average rating for each movie, average rating of each user and a weighted average of the two to estimate linear regression weights
- Matrix factorization with stochastic gradient descent algorithm was implemented with different parameters' setting such as learning rate, penalty and number features. In addition, matrix factorization using alternating least squares was implemented
- The Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) over 5 folds of cross-validation were used as the evaluation metrics
- The project was implemented in Python

Spooky Authors Identification

PROBLEM: PREDICT THE AUTHOR OF EXCERPTS FROM HORROR STORIES USING NLP TECHNIQUES (KAGGLE CHALLENGE)

- Features extraction from text was implemented using normalized Term Frequency - Inverse Document Frequency (TF-IDF) weights
- Multinomial Naive Bayes (MNB) and Logistic Regression (LR), along with Cross Validation (CV) were implemented to choose the best number of features and tuning parameters for both models
- The evaluation metric was the Categorical Cross Entropy (CCE)
- The results obtained were among the top 50% on Kaggle evaluation
- The project was implemented in Python

Predicting Ratings of Amazon Products Using Review Texts (Best project award)

PROBLEM: IS IT POSSIBLE TO PREDICT THE RATING (1-5) OF A PRODUCT BASED ON CONSUMER'S OPINION (TEXT)?

- NLP techniques were used to preprocess 82M reviews in a parallel manner using Dask, a Python library for distributed and parallel computing
- Preprocessed reviews were converted to Term Frequency - Inverse Document Frequency (TF-IDF) in a distributed fashion using the Distributed ASCII Supercomputer 3 (DAS-3) of Leiden University's Data Science Lab
- The models used are Linear Regression, Logistic Regression, Naive Bayes, and Multilayer Perceptron (MLP) Neural Network
- All models were evaluated using the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE).
- This project was implemented in Python.

Optimizing Optical Layers Thickness

PROBLEM: FIND THE OPTIMAL THICKNESS OF 30 OPTICAL LAYERS

- Evolution Strategy was applied to find the optimal thickness of layers that allows filtering certain wavelength with a fixed evaluation budget
- Different variations of Evolution Strategy was implemented such as different ways of mutation and recombination
- Each variations' results were averaged over 30 independent runs and compared.
- The project was implemented in Matlab

Denoising Dirty Documents

PROBLEM: REMOVE NOISE FROM "DIRTY" DOCUMENTS (IMAGES)

- Convolutional Autoencoders were used to learn the cleaned representation of a scanned document
- A variety of experiments were conducted using different number of layers, kernel sizes and hyperparameters
- The project was implemented using Keras in Python

Additional Activities

European Innovation Academy

Nice, France

PARTICIPANT

July 2014

- Served as the technical lead and created a minimum viable product from an idea in 14 days
- Pitched to a panel of 10 investors and 200+ people in request of incubation and funding

Entrepreneurship Innovation Trip

Silicon Valley, CA, USA

PARTICIPANT

March 2014

- Visited various tech companies and start ups such as Twitter, Whatsapp, YouNoodle and oDesk
- Interacted with entrepreneurs and investors in an open panel discussion

Carnegie Mellon Qatar Men's Basketball Team

Doha, Qatar

SHOOTING GUARD

Sept. 2012 - May 2016

- Served as the co-captain, and was responsible for leading team practice
- Named Most Valuable Player two consecutive years