

Ahmed Ktob



Data Scientist

Address 103 Bld. 4, SanJianFang, Chaoyang
Beijing, 100025
China

Email ktobah@gmail.com

Phone +86 13426054970

Links [Github](#), [LinkedIn](#)

01 PROFILE

A detail-oriented professional with 2+ years of experience and proven knowledge of machine learning, data analysis, and predictive modeling. I aim to successfully leverage my skills to fill a data scientist/software engineer role.

02 EMPLOYMENT HISTORY

03/2019 — Present

Data Scientist at ControlExpert

Beijing

- Identify valuable data sources and automate collection processes.
- Analyze large amounts of information to discover trends and patterns.
- Undertake the pre-processing of structured and unstructured data.
- Build car damage detection, segmentation, and decision aggregation models using ML/DL techniques.
- Perform extensive hyper-parameters tuning.
- Combine models through ensemble modeling.
- Explore structured data & implementation of NLP algorithms.
- Develop auto labeling tools.
- Report findings through data visualization techniques.
- Collaborate with engineering and product development teams.
- Propose solutions and strategies to business challenges.

09/2013 — 06/2018

Researcher at Beihang University

Beijing

I worked at Prof. Zhoujun Li's lab as a researcher on semantic web and knowledge graphs. My research project was on the construction of an Arabic Knowledge Graph. I identified the challenges facing its creation and the opportunities it offers. We were able to create a small knowledge graph and built two tools that would allow practitioners and data owners to transform their data into RDF and share it with the world. Also, this research project led to three peer-reviewed publications. The applications of the AKG are beyond description; it can be used in creating semantic-based Arabic search engines, build smarter Arabic AI applications, enhance question-answering systems, and much more.

03 INTERNSHIPS

Oct 2015 — Apr 2016

Software Test Engineer at Symbio

Beijing

- Design test plans, scenarios, scripts, or procedures.
- Document software defects, using a bug tracking system, and report defects to software developers.
- Identify, analyze, and document problems with program function, output, UI, or content.
- Monitor bug resolution efforts and track successes (using Jira).
- Provide feedback and recommendations to developers on software usability and functionality.
- Conduct software compatibility tests with hardware and operating systems.

04 EDUCATION

Sep 2014 — Jun 2018

Beihang University

China

PhD

Major: Computer Software and Theory

Jan 2018 — May 2018

University of Liege

Belgium

Exchange Program

Sep 2013 — Jun 2014

Beihang University

China

Certificate

Chinese Language Training Course

Sep 2011 — Jul 2013

Djillali Liabes University

Algeria

Master

Major: Networks, Systems and Information Security

Sep 2008 — Jul 2011

Dr. Tahar Moulay University

Algeria

Bachelor

Major: Computer Science

05 SKILLS

Research / Knowledge Graph

Pandas / Numpy / Seaborn

ML / Deep Learning

Linux / APIs / Docker / AWS

Data Collection / Cleansing

Collaboration & Teamwork

Data Analysis / Visualization

Communication / Creative Thinking /
Active Learning

Python / Java

Predictive Modeling / Decision Analytics

TensorFlow / Keras / SciKit

06 LANGUAGES

Arabic



French



English



Chinese



07 PROJECTS

Jan 2021 — Present

Structured Data & Graph Database

ControlExpert

A highly-collaborative project that aims to enhance CE's benefit from the gathered structured data. I work with the team in analyzing the data and extracting a data model to build a graph database using Neo4j. This database will allow CE to explore the hidden connections between claims, car drivers, reporters, investigators, and so on to eventually discover frauds.

Dec 2020 — Present

Damage & Parts Segmentation

ControlExpert

This project's goal is to solve an issue with part detection that prevents us from delivering accurate results on the damaged parts. Also, we aim to quantify the damage so I started working on both car damage and parts segmentation. The initial results seem promising.

Nov 2019 — Present

Damage Detection

ControlExpert

Lead the development of a brand-new model to detect the damaged parts in a car accident. This predictive model will be integrated into our risk control service to improve our accuracy. The first step to achieving such a goal is data. Thus, I collect data, check it for completeness, clean it, upload images to our AWS cloud, and update the labeling MongoDB database. I oversee the labeling and guarantee that quality is met. Secondly, I use techniques of transfer learning to train a variety of CNN models. Tools such as Tensorflow, Keras, and Docker are used for model development and training, and Wandb for logging metrics. Afterward, an analysis of the results is conducted, and results are shared with the team in the form of detailed reports or brief summaries in Microsoft Teams.

Aggregation & Thresholds

Technically, decisions are taken around the nuances of thresholds. Therefore, a well-defined threshold can make a significant impact. In this project, I continuously work on optimizing our threshold for better decision making. The idea is that a decision taken from several images is better than from one. Based on this concept, I use Keras, with Tensorflow as the backend, to create an aggregation model that suits our needs. I build different architectures, sequential with dense or convolutional layers and functional multi-branch, then use Optuna to run an extensive (over 6000 experiments) hyper-parameters search. An innovation here was the use of convolutional layers with structured data (numbers) which proved useful. Once training is finished, results are evaluated, and top models are picked for further threshold optimization. This project gave us a 10-15% accuracy boost.

AutoLabeler

At CE, we recognized the bottleneck of slow data labeling. Thus, I worked on creating a labeling tool that used one of our internal object detection models to automatically label our images. Further, human labelers would check if the results are accurate and amend them if necessary. This tool had an accuracy of 80% and allowed us to, theoretically, save over 400 hours of human labor.

MongoDB Management

Our labeling tool uses a MongoDB database to store data. I am responsible for keeping the database up to date, add and remove data, fix any issues that may arise, and automate the generation of labeling stats for our human experts. Also, I create scripts that are used by the team to generate data for model training. In addition, I was one of two members responsible for checking the quality of our human experts until August 2019 where a new colleague took over this role.

08 PUBLICATIONS

Ahmed Ktob and Zhoujun Li, "Marefa: Turning Publisher's Catalog Data into Linked Data". doi: 10.4018/IJSWIS.2018070108. (refereed paper)

Ahmed Ktob and Zhoujun Li, "The Arabic Knowledge Graph: Opportunities & Challenges". doi: 10.1109/ICSC.2017.22 (Refereed paper)

Ahmed Ktob, Zhoujun Li and Djelloul Bouchiha, "Extracting Linked Data from HTML Tables". doi: 10.1109/CIC.2017.00018 (Refereed paper)

Mohamed Abdillah, Li Zhang, Jing Jiang and **Ahmed Ktob**, "Predicting Which Pull Requests Will Get Reopened in GitHub". doi: 10.1109/APSEC.2018.00052. (Refereed paper)

09 COURSES

DataQuest

- Python for Data Science: Intermediate (Jan 2019)
- Python for Data Science: Fundamentals (Jan 2019)
- Statistics Fundamentals (Jan 2019)
- APIs and Web Scrapping (Dec 2019)
- Command Line: Beginner (Dec 2018)
- Command Line: Intermediate (Dec 2018)
- Git and Version Control (Dec 2018)
- SQL and Databases: Advanced (Dec 2018)
- SQL Intermediate: Table Relations and Joins (Dec 2018)
- SQL Fundamentals (Dec 2018)
- Data Cleaning (Nov 2018)
- Exploratory Data Visualization (Nov 2018)
- Pandas and Numpy Fundamentals (Nov 2018)
- Python Programming: Intermediate (Nov 2018)
- Python Programming: Beginner (Nov 2018)
- Storytelling Through Data Visualization (Nov 2018)

Coursera

- DeepLearning.AI TensorFlow Developer Professional Certificate (Dec 2020)
- IBM AI Foundations for Business Specialization (Oct 2020)
- Data Science Orientation (Sep 2020)
- Python Programming: A Concise Introduction (Jul 2017)
- Introduction to HTML (Sep 2016)
- Introduction to Meteor.js Development (Sep 2016)
- Agile Meets Design Thinking (Aug 2016)
- Front-End Web UI Frameworks and Tools (Apr 2016)
- Hadoop Platform and Application Framework (Apr 2016)
- HTML, CSS and JavaScript (Mar 2016)
- Grammar and Punctuation (Mar 2016)
- Introduction to Big Data (Feb 2016)
- Learn to Program: The Fundamentals (Nov 2012)
- Computer Science 101 (Jun 2012)

10 HONORS & AWARDS

Pluralsight Python Core Language Test - Proficient 192 (Oct 2020)
TestDome General and Python Data Science Test - Top 25% (Sep 2020)
ControlExpert's Star of the Year (Jan 2020, Beijing)
ControlExpert's Star of the Month (Oct 2019, Beijing)
Postdoc: Swiss Government Excellence Scholarship (May 2018, Zurich)
Distinguished Foreign Students Scholarship of Beijing Government and Beihang University of Excellent Study (Dec 2017, Beijing)
Beihang YuanHang Scholar (Sep 2017, Beijing)
Cisco Global Cybersecurity Scholarship (Jun 2017, Online)
Chinese Council Scholarship (Sep 2013 - Jun 2018, Beijing)
Algerian Ministry of Higher Education Scholarship (Sep 2013 - Jun 2018, Beijing)
Master First Class Honor (Jul 2013, Saida)
CCNA Exploration: Network Fundamentals (Mar 2012, Saida)
Bachelor Second Class Honor (Jul 2011, Saida)

11 HOBBIES

Football, reading, walks in nature, and travelling.

12 EXTRA-CURRICULAR ACTIVITIES

Apr 2014 — Present

Fundraising, Beihang University

Beijing

Since 2014 and until July 2018, I was the chairman and founder of a student-created charity group within Beihang University where on the course of three years we collected over CNY 70000. Currently, I am still an active member of this operational group but under different leadership.

13 MEMBERSHIPS

Australian Computer Society (ACS) Member (2020 - 2021)
ACM Student Member (2016 - 2018)
IEEE Student Member (2016 - 2018)

14 REFERENCES

References available upon request.