

# Shahab Helmi

[shahab.helmi@ucdenver.edu](mailto:shahab.helmi@ucdenver.edu) | <http://www.shelmi.com> | <https://www.linkedin.com/in/shelmi>

## SUMMARY

I have been developing custom applications using .NET, Java, and C++ for the past 9 years as well as designing data science algorithms for that past 4 years. My research interests are based on Frequent Pattern Mining from Spatiotemporal datasets and I am very passionate about software development with a focus on machine learning.

I put a lot of emphasis on details as it is shown through my deep understanding of technology while I strive to get to the root of problems as they occur. I have strong communication skills and enjoy working with, helping, and learning from my colleagues.

I am looking to fill a position as a software developer or data scientist where I can write elegant codes and design novel ML algorithms to solve complex, real world problems.

## EDUCATION

### Ph.D. in Computer Science

University of Colorado Denver, CO

*Expected: December 2019*

### Master of Science in Computer Science

University of Southern California, Los Angeles, CA

*December 2014*

### Bachelor of Science in Software Engineering

Azad University of Tehran, Tehran, Iran

*June 2012*

## EXPERIENCE

### PROFESSIONAL EXPERIENCES

#### Data Scientist Intern

TransUnion, Greenwood Village, CO

*Since May 2018*

- As a data scientist intern, I have been involved in various tasks, such as improving existing models, building new models, developing tools for data analysis, QA, helping clients improve their estimation tools by analyzing their data and finding data-related issues.
- **Technologies:** R, C#, SQL, Microsoft Excel, Beyond Compare, Birst

#### COPD Date Warehouse and Data Mining

University of Colorado Denver, Anschutz Medical/Downtown Campus, National Jewish Health, Denver, CO

*Since January 2017*

- This project, which is sponsored by the National Institutes of Health (NIH), is two-fold. For the first component of this project, we developed tools for efficient data integration, storage, query, and analysis for the effective data-driven study of COPD. As for the second component, we are currently applying various high-dimensional clustering methods, such as subspace and projective clustering algorithms, to the COPDGene datasets. For more information, please visit the BDLab website at <http://cse.ucdenver.edu/~bdlab/research.html>.
- **Technologies:** C#, WPF, LINQ, SQL, R, Python

#### Synthea-OMOP ETL Tool

University of Colorado Denver, Anschutz Medical Campus, Denver, CO

*June 2017 - September 2017*

- Developed an ETL tool that extracts data from [Synthea FHIR](#) outputs, transforms it to the [OMOP](#) common data model, and loads it to an OMOP-based database hosted on the PostgreSQL server. This application also handles missing data, missing vocabulary, etc.
- **Technologies:** Java, JSON, PostgreSQL, IntelliJ

#### RadOn

University of Colorado Denver, Anschutz Medical Campus, Denver, CO

*September 2016 - September 2017*

- Developed an application for the Radiation Oncology Department that pulls treatment data from different sources, such as prescriptions provided by different physicians, treatment plans created by dosimetrists, and information of actual treatments carried out by the treatment instruments. The application then processed

the retrieved data, identifying inconsistencies among the sources to ensure that the treatment was going according to plan. For more information, please visit the [BDLab](#) website.

- **Technologies:** C#, WPF, Microsoft SQL Server, LINQ, Eclipse Scripting API

### **CiSoft**

*January 2014 - September 2014*

University of Southern California, Los Angeles, CA

- Implemented an application using Microsoft StreamInsight to receive streaming sensor data from Chevron to detect and reconstruct errors.

### **iWatch**

*May 2013 - December 2013*

University of Colorado Denver, Anschutz Medical Campus, Denver, CO

- This project was implemented for the Department of Public Safety at the University of Southern California to detect and announce crimes based on Geospatial data.
- **Technologies:** Oracle Spatial Database, Microsoft SQL Server, Java, JDBC, and Microsoft C#.NET

## **TEACHING**

### **Instructor/Teaching Assistant**

*January 2014 – December 2017*

University of Southern California and University of Colorado Denver

- **Courses:** Big Data Systems, Database System Concepts, C++ Programming, Database System Concept, Introduction to Data Science, Introduction to Data Mining
- **Responsibilities:** designing the curriculum, creating and presenting lectures, grading assignments

## **SKILLS**

- **Programming Languages:** C#, Java, Python, C++, R
- **Cloud:** Azure ML
- **Scripting Languages:** SQL [Microsoft SQL Server, PostgreSQL, MySQL, Oracle], JavaScript, HTML
- **Data Science and Machine Learning:** Scikit-learn, TensorFlow, CRAN
- **Data Processing and Visualization:** git, Tableau, Microsoft Power-BI, Matplotlib, Microsoft Excel
- **Big Data:** Apache Spark, Apache Kafka

## **CERTIFICATES**

- **Microsoft Professional Program for Data Science:** Query Relational Databases, Introduction to Python for Data Science, Essential Statistics for Data Analysis using Excel
- **Udemy Certificates:** Complete Python Bootcamp, Tableau 9 For Data Science, R Programming A-Z™

## **PUBLICATIONS**

- S. Helmi, F. Banaei-Kashani, "Spatiotemporal Range Pattern Queries on Large-scale Co-Movement Pattern Datasets", in IEEE International Conference on Big Data, 2017
- S. Helmi, F. Banaei-Kashani, "Efficient Processing of Spatiotemporal Pattern Queries on Historical Frequent Co-Movement Pattern Datasets", in proceedings of the 43rd VLDB International Conference on Mobility Analytics for Spatio-temporal and Social Data, 2017
- S. Helmi, F. Banaei-Kashani, "Mining Frequent Episodes from Multivariate Spatiotemporal Event Sequences", in proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming. ACM, 2016.

## **AWARDS**

- [Student Travel Award](#), 2017 IEEE International Conference on Big Data, Boston, MA
- Won the 1<sup>st</sup> place in the Transunion Healthcare Hackathon, May 2018 and Nov 2018

## **COMMUNITY SERVICES**

- **Research Paper Peer Reviewer:** BSD2018, CSAE 2018, PMC 2016, ACM TSAS 2016, ACM SigSpatial 2015
- **Webmaster:** [cse.ucdenver.edu/~BSD2018/](http://cse.ucdenver.edu/~BSD2018/), [cse.ucdenver.edu/~bdlab/research.html](http://cse.ucdenver.edu/~bdlab/research.html)