

Cagri Ozcaglar

Sr. Machine Learning Engineer
LinkedIn
Sunnyvale, CA

cagriozcaglar@gmail.com
<http://www.cs.rpi.edu/~ozcagc2>
<https://github.com/cagriozcaglar>

- Skills**
- **Languages:** Java, Python, Scala, C++, R, MATLAB, Shell Scripting, Groovy, Perl, PHP, HTML.
 - **Modeling Software:** Tensorflow, Keras, Python (scikit-learn), R, MATLAB, Amazon Machine Learning, Photon-ML.
 - **Big Data Processing:** Spark, Hadoop, Apache Pig, AWS.
 - **Database Management:** SQL, PostgreSQL, MySQL, Oracle 10g/11g.
 - **Tools and Applications:** Latex, Eclipse, IntelliJ, Visual Studio .NET, Git, SVN.
- Research Interests** Search and Recommendation Systems, Machine Learning, Data Mining, Data Fusion, User Intent Modeling, Social Network Analysis, Multiway Analysis, Bioinformatics.
- Education**
- **Ph.D., Computer Science** 2008-2012
Rensselaer Polytechnic Institute, Troy, NY
Thesis: Algorithmic Data Fusion Methods for Tuberculosis
 - **M.S., Computer Science** 2006-2008
Rensselaer Polytechnic Institute, Troy, NY
Thesis: Classification of Email Messages into Topics Using Latent Dirichlet Allocation
 - **B.S., Computer Science** 2002-2006
Bilkent University, Ankara, Turkey
- Experience**
- **Sr. Machine Learning Engineer, LinkedIn** Sunnyvale, CA
Sr. Machine Learning Engineer 2018 - Present
Teams: Careers Relevance, Down-Funnel Optimization
 - Worked on Job Recommendation system ranking component migration from GLMix model to deep learning models using Tensorflow.
 - Led, designed, and implemented Unified Offline Ranking EvaluatiOn (OREO) for Recruiter and Jobs Ecosystem of LinkedIn.
 - Worked on all aspects of Job Search Relevance, including query recommendations, candidate selection, and ranking.
 - Led Guided Search project for making query recommendations for Search starters and Inline query suggestions. Filed one patent.
 - Designed Unified Guided Search flow for Guided Search. Implemented and productionized Seq2Seq-LSTM model for generating query suggestions with online inference on free-form queries in Search starter and Inline Query Suggestion use cases.
 - Improved candidate selection by adding high-quality skill-ID-based retrieval and skill-to-title expansion. Built a rule learning model to generate rule precedence, translated to Galene (based on Lucene), which improved candidate selection significantly.
 - Improved Job Search ranking with various changes: 1) Labeling strategy update (Premium / Basic job applies are equal-labeled), 2) Feature Engineering using Frame: Added search-job skill-coverage percentage features to be used in ranking, 3) Using member / job skill / title topic embeddings in ranking.

- Topic Embeddings for Job Search: Designed the use of topic embeddings for Job Search, implemented and productionized topic embedding match features used in L2 ranking.

Machine Learning Engineer

2016 - 2018

Team: Talent Relevance

- Led a team of software engineers and researchers for the new Contextual Search product. Designed the offline and online architecture for Top-N skill recommendations for Contextual Search, and contributed to the charter launch in May 2018.
 - Improved Recruiter Search ranking using Contextual skill match features and topic vector match features via Latent Dirichlet Allocation. Designed and implemented offline and online pipeline for generating these features.
 - Designed, implemented, and productionized Generalized Linear Mixed (GLMix) models for Recruiter Search ranking with Learning-to-Rank features, XGBoost model scores and tree interaction features. This model led to 8.5% / 5% / 2% lift in InMail Accept@1/@5/@25 respectively.
 - Designed and built Generalized Linear Mixed (GLMix) models for improving Recruiter search relevance. Improved InMail Accept@k by 10%. Filed one patent.
 - Contributed to the design and implementation of an end-to-end deep learning model building pipeline for Recruiter Search using Tensorflow, Keras, Spark, Scala.
 - Designed and implemented an end-to-end Offline Feature Engineering Workflow which allows users to quantify the impact of adding new features to Recruiter Search Modeling pipeline with hypothesis tests on the key evaluation metric.
 - Designed and implemented a Question Answering / Preference Elicitation system for Job Posting Flow, with adaptive question selection using entropy maximization after each question.
 - Conducted data analysis and built tree models to identify member skill features that distinguish confirmed hires from only-InMail-accepters for given (company, title) pairs, for Contextual Search product.
 - Contributed to the migration of Recruiter Search label generation pipeline from Hadoop / Pig stack to Spark / Scala stack.
 - Published 3 articles ([SIGIR'2018](#), [CIKM'2018](#), [WWW'2019](#)), 10 patents, and 1 blog post on [The AI Behind LinkedIn Recruiter Search and Recommendation Systems](#).
- **Research Scientist, Amazon** Seattle, WA
Research Scientist 2013 - 2016
Team: Consumer Marketing Analytics
 - Designed, built, and productionized product category purchase propensity models within Amazon in 11 marketplaces using weighted logistic regression, including General Ledger Purchase (GLP) and Browse Node Purchase (BNP) propensity models.
 - Designed, built, and productionized channel response propensity models using weighted logistic regression, including Email Risk and Response models (ERR), Mobile click and conversion propensity models.
 - Designed, built, and productionized uplift / incremental models to calculate the differential propensity of a customer to take an action after targeting. Various applications include direct mail targeting for streaming Prime Instant Video (PIV), making a purchase from Fashion product categories, signing up for Amazon Business program. Live test results for PIV streaming show that incremental models return a lift of 32% and 121% on the percentage of streamers respectively, compared to overall response models and random targeting.
 - Designed and built uplift models with importance weighting, in order to measure the differential propensity of a customer to take an action after targeting in case of treatment set selection bias. In live experiments, bias-corrected uplift models returned higher incremental response rate compared to treatment model and biased uplift model.

- Mobile channel acquisition and engagement models: Designed, built, and productionized Amazon Mobile shopping app download propensity models for mobile acquisition. Designed, built, and productionized mobile channel adjusted product category purchase propensity models for engagement with mobile channel. In live tests, compared to Mobile App First Sign-in (FSI) rate of 0.3% for universal control set, targeting Mobile shopping app download propensity model based segments with BAU offer and 5\$ incentive offer returned 413% and 1476% incremental lift on FSI rate, respectively.
- Designed, built, and productionized Prime Free-Trial program sign-up propensity models using thresholded model ensembles. Using two different ranking methods in live experiments, customer segments selected based on Prime Free-Trial sign-up propensity models returned 94% and 74% higher Prime Free-Trial program sign-up rates compared to the baseline sign-up rate of 7%.
- Designed and ran live experiments for targeting Amazon customers on Facebook, using browse node purchase propensity models (BNP), and deal seeker propensity models. In live tests, campaigns targeting browse node purchase propensity segments drove \$431K at 23.4% E%O, and campaigns targeting deal seeker segments along with BNP segments drove \$15MM at 14.67%.
- Designed, built, and productionized various channel usage propensity models and customer life-cycle propensity models, including Kindle Cross-Platform reader download propensity models (KCP), Google channel reliance models, Amazon Attrition propensity models.
- Helped internal customers onboard to Predix, an automated predictive modeling platform, by helping them with model target generation, model building, and model evaluation.
- Organized and held Machine Learning Talk Series for Consumer Marketing organization as a monthly recurring event.
- Published articles, presented posters, gave talks in Amazon internal machine learning conferences.
- Conducted 50+ interviews for hiring research scientists to various teams.
- **Software Developer, Bank of America Merrill Lynch** New York, NY
Software Developer 2012 - 2013
Team: Equity Linked Technology
 - Designed and implemented software for processing TESS real time feeds within trading systems.
 - Designed and implemented a connector between Access Request Management (ARM) and RAM database.
- **Rensselaer Polytechnic Institute** Troy, NY
Research Assistant 2006 - 2012
 - **TB-Insight:** <http://tbinsight.cs.rpi.edu>
Supervisor: Prof. Bulent Yener (PI: Prof. Kristin Bennett)
 - UBF: Developed the Unified Biclustering Framework (UBF) to find host-pathogen associations among *M. tuberculosis* complex strains and TB patients.
 - Developed an algorithm to find the mutation history in the DR region of *M. tuberculosis* complex and found topological attributes of the resulting phylogenetic tree. Built two new models for mutation length frequency: Starting Point Model (SPM) and Longest Block Model (LBM).
 - TCF: Developed the Tensor Clustering Framework (TCF) to cluster *M. tuberculosis* complex strains into coherent groups using multiple-biomarker tensors.

- TB-Vis: Designed and implemented a visualization program for pathogen and host analysis of tuberculosis. The visualizations include: 1) Spoligo forests which display spoligotype evolution using various distance measures of genomic data of *M. tuberculosis* complex, 2) Host-pathogen maps which display patient characteristics classified by the genotype of *M. tuberculosis* complex which infects the patients. Spoligo forests are used in TB-Lineage tool at http://tbinsight.cs.rpi.edu/about_tb_lineage.html . Both spoligo forests and host-pathogen maps are used in the design of interactive version of TB-Vis at http://tbinsight.cs.rpi.edu/about_tb_vis.html .

- **Enron Email Classification Into Topics**

Supervisor: Prof. Sibel Adali, Prof. Boleslaw Szymanski

- Classified email messages of Enron into topics using the Latent Dirichlet Allocation. Described new metrics for classification assessment of email distribution into topics.

- **MetPetDB:** <http://metpetdb.rpi.edu>

Supervisor: Prof. Sibel Adali, Prof. Boleslaw Szymanski

- Contributed to the design and implementation of the client-server architecture of a database for Metamorphic Petrology.

- **TUBITAK (Scientific and Technological Research Council)**

Ankara, Turkey

Software Engineer Intern

Summer 2005

- Designed and implemented a program which calculates the salaries of the employees using their entrance and exit time to/from the building.

- **TRT (Turkish Radio and Television)**

Ankara, Turkey

Software Engineer Intern

Summer 2004

- Contributed to the implementation of user interface of Eurovision Song Contest website.

Teaching

- **Teaching Assistant**

Rensselaer Polytechnic Institute

- Introduction to Algorithms Fall 2010
- Computer Science II: Data Structures Spring 2008
- Computer Organization Spring 2007
- Artificial Intelligence Fall 2006

- **Undergraduate Teaching Assistant**

Bilkent University

- Discrete Mathematics Spring 2006

Journal Articles

- S. Salem, C. Ozcaglar. **Hybrid coexpression link similarity graph clustering for mining biological modules from multiple gene expression datasets.** *BioData Mining*, 2014.
- K. P. Bennett, C. Ozcaglar, J. Ranganathan, S. Raghavan, J. Katz, D. Croft, B. Yener, A. Shabbeer. **TB-vis: Visualizing TB patient-pathogen relationships.** *Tuberculosis*, 2013.
- C. Ozcaglar, A. Shabbeer, N. Kurepina, N. Rastogi, B. Yener, K. P. Bennett. **Inferred spoligo forest topology unravels spatially bimodal distribution of mutations in the DR region.** *IEEE Transactions on NanoBioscience*, 2012.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Epidemiological models of *Mycobacterium tuberculosis* complex infections.** *Mathematical Biosciences*, 2012. (Featured as the most downloaded paper of *Mathematical Biosciences* journal in March-May 2012).

- A. Shabbeer, L. Cowan, C. Ozcaglar, N. Rastogi, S. L. Vandenberg, B. Yener, K. P. Bennett. **TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex.** *Infection, Genetics and Evolution*, 2012.
- A. Shabbeer, C. Ozcaglar, B. Yener, K. P. Bennett. **Web tools for molecular epidemiology of tuberculosis.** *Infection, Genetics and Evolution*, 2012. (Featured as the most downloaded paper of *Infection, Genetics and Evolution* journal as of December 2011).
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Sublineage structure analysis of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors.** *BMC Genomics*, 2011.

**Conference/
Workshop
Articles**

- C. Ozcaglar, S. Geyik, B. Schmitz, P. Sharma, A. Shelkovnykov, Y. Ma, E. Buchanan. **Entity Personalized Talent Search Models with Tree Interaction Features.** *WWW*, 2019.
- R. Ramanath, H. Inan, G. Polatkan, B. Hu, Q. Guo, C. Ozcaglar, R. Wu, K. Kenthapadi, S. Geyik. **Towards Deep and Representation Learning for Talent Search at LinkedIn.** *CIKM*, 2018.
- S. Geyik, Q. Guo, B. Hu, C. Ozcaglar, R. Wu, K. Kenthapadi. **Talent Search and Recommendation Systems at LinkedIn: Practical Challenges and Lessons Learned.** *SIGIR*, 2018.
- S. Salem, C. Ozcaglar. **MFMS: Maximal frequent module set mining from multiple human gene expression datasets.** *KDD International Workshop on Data Mining in Bioinformatics (BIOKDD)*, Chicago, August 2013.
- C. Ozcaglar, A. Shabbeer, N. Kurepina, B. Yener, K. P. Bennett. **Data-driven insights into deletions of *Mycobacterium tuberculosis* complex chromosomal DR region using spoligoforests.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Atlanta, November 2011.
- K. P. Bennett, C. Ozcaglar, J. Ranganathan, S. Raghavan, J. Katz, D. Croft, B. Yener, A. Shabbeer. **Visualization of tuberculosis patient and *Mycobacterium tuberculosis* complex genotype data via host-pathogen maps.** *IEEE BIBM Workshop on Computational Advances in Molecular Epidemiology*, Atlanta, November 2011.
- M. Aminian, A. Shabbeer, K. Hadley, C. Ozcaglar, S. Vandenberg, K. P. Bennett. **Knowledge-based Bayesian network for the classification of *Mycobacterium tuberculosis* complex sublineages** *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB)*, Chicago, August 2011.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Examining the sublineage structure of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Hong Kong, December 2010.
- A. Shabbeer, C. Ozcaglar, M. Gonzalez, K. P. Bennett. **Optimal Embedding of Heterogeneous Graph Data with Edge Crossing Constraints.** *NIPS Workshop on Challenges of Data Visualization*, Whistler, BC, Canada, December 2010.
- J. M. Pyle, F. S. Spear, S. Adali, B. K. Szymanski, S. Pearce, A. Waters, Z. Linder, C. Ozcaglar. **MetPetDB: The Unique Aspects of Metamorphic Geochemical Data and Their Influence on Data Model, User Interface and Collaborations.** *Geological Society of America Abstracts with Programs*, Vol. 39, No. 6, 2007.

**Technical
Reports**

- A. Shabbeer, C. Ozcaglar, K. P. Bennett. **Crossing minimization within graph embeddings.** *arXiv*, 2012.
- C. Ozcaglar, B. Yener, K. P. Bennett. **Host-pathogen association analysis of tuberculosis patients via Unified Biclustering Framework.** Rensselaer Polytechnic Institute. TR-12-05, 2012.

- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors.** Rensselaer Polytechnic Institute. TR-10-08, 2010.
 - F. S. Spear, J. M. Pyle, S. Adali, B. K. Szymanski, A. Waters, Z. Linder, C. Ozcaglar, S. O. Pearce. **MetPetDB: A database for metamorphic geochemistry.** Rensselaer Polytechnic Institute. TR-08-14, 2008.
- Patents**
- C. Ozcaglar, R. Ranjan, V. Parimi. **Uplift Modeling with Importance Weighting.** *Issued*, June 2020.
 - C. Ozcaglar, K. Kenthapadi. **Candidate Team Recommendations.** *Filed*, November 2019.
 - C. Ozcaglar, K. Kenthapadi. **Detecting Anomalous Candidate Recommendations.** *Filed*, November 2019.
 - K. Kenthapadi, C. Ozcaglar. **Feedback-based Update of Candidate Recommendations.** *Filed*, November 2019.
 - C. Ozcaglar, G. Borje, S. Geyik, G. Gulati, K. Thakkar. **Contextual Search Ranking Using Entity Topic Representations.** *Filed*, June 2019.
 - S. Wakankar, M. Meng, C. Ozcaglar, V. Abdrashitov. **Data Selection Based on Career Transition Embeddings.** *Filed*, June 2019.
 - C. Ozcaglar, S. Geyik, P. Sharma, B. Schmitz, E. Buchanan. **Entity-Level Search Models with Tree Interaction Features.** *Filed*, August 2018.
 - P. Cheung, E. Buchanan, C. Liao, D. Boyd, G. Gulati, F. Or, C. Ozcaglar. **Utilizing Search Facets Based on Project Context.** *Filed*, July 2018.
 - R. Ramanath, G. Polatkan, Q. Guo, C. Ozcaglar, K. Kenthapadi, S. Geyik. **Unsupervised Learning of Entity Representations Using Graphs.** *Filed*, July 2018.
 - R. Ramanath, G. Polatkan, Q. Guo, C. Ozcaglar, K. Kenthapadi, S. Geyik. **Generating Supervised Embedding Representations for Search.** *Filed*, July 2018.
 - R. Ramanath, G. Polatkan, Q. Guo, C. Ozcaglar, K. Kenthapadi, S. Geyik. **Generating Supervised Embeddings Using Unsupervised Embeddings.** *Filed*, July 2018.
 - R. Ramanath, G. Polatkan, Q. Guo, C. Ozcaglar, K. Kenthapadi, S. Geyik. **Generating Candidates for Search Using Scoring / Retrieval Architecture.** *Filed*, July 2018.
 - R. Ramanath, G. Polatkan, Q. Guo, C. Ozcaglar, K. Kenthapadi, S. Geyik. **Applying Learning-to-Rank For Search.** *Filed*, July 2018.
 - C. Ozcaglar, V. Dialani, S. Smoot, S. Geyik, A. Nair. **Feature Selection Impact Analysis for Statistical Models.** *Filed*, December 2017.
 - C. Ozcaglar, R. Wu, J. Yang, A. Gupta, A. Nair. **Generalized Linear Mixed Models (GLMix) for Improving Search.** *Filed*, December 2017.
- Theses**
- **Algorithmic Data Fusion Methods for Tuberculosis,** *Ph.D. thesis*, Rensselaer Polytechnic Institute, 2012.
 - **Classification of Email Messages into Topics Using Latent Dirichlet Allocation,** *M.S. thesis*, Rensselaer Polytechnic Institute, 2008.
- Poster Presentations**
- C. Ozcaglar, B. Yener, K. P. Bennett. **UBF: Unified Biclustering Framework.** New York Academy of Sciences (NYAS) 7th Annual Machine Learning Symposium, NYC, October 2012.
 - C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors.** *RPI-NSF Workshop on Multiscale Modeling of Complex Data*, Troy, NY, September 2011.

- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Insights into Camin-Sokal Parsimony and evolution of spoligotypes via spoligoforests.** *New York Academy of Science (NYAS) Symposium on Imaging, Visualization and Simulation: New Tools for Technology and Healthcare*, NYC, June 2011.
- C. Ozcaglar, A. Shabbeer, S. Vandenberg, B. Yener, K. P. Bennett. **Multiple-biomarker tensor analysis for tuberculosis lineage identification.** *NIPS Workshop on Tensors, Kernels and Machine Learning*, Whistler, BC, Canada, December 2010.
- C. Ozcaglar, B. Yener, A. Shabbeer, M. Aminian, K. P. Bennett. **A clustering framework for *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors.** *New York Academy of Science (NYAS) 5th Annual Machine Learning Symposium*, NYC, October 2010.
- C. Ozcaglar, B. Yener, A. Shabbeer, M. Aminian, K. P. Bennett. **Examining sublineage structure of *Mycobacterium tuberculosis* complex strains with multiway modeling.** *Eigenvector University*, Seattle Washington, May 2010. (Best poster award)

Talks

- **MFMS: Maximal frequent module set mining from multiple human gene expression datasets.** *ACM SIGKDD International Workshop on Data Mining in Bioinformatics (BIOKDD)*, Chicago, August 2013.
- **TCF: Tensor clustering framework on multiple-biomarker tensors.** *Bogazici University, Middle East Technical University, Bilkent University*, January 2012.
- **Data-driven insights into deletions of *Mycobacterium tuberculosis* complex chromosomal DR region using spoligoforests.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Atlanta, November 2011.
- **Examining the sublineage structure of *Mycobacterium tuberculosis* complex strains with multiple-biomarker tensors.** *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Hong Kong, December 2010.
- **Extracting Associations from Activities.** *NSF Cyber-Enabled Discovery and Innovation Symposium*, Troy, NY, September 2007.

Awards

- Transaction Risk Management Systems Modeling Hackathon at Amazon, first place, 2015.
- Student Travel Award to attend IEEE BIBM 2010.
- Full scholarship awarded by Rensselaer Polytechnic Institute for graduate study, 2006 - 2012.
- Full scholarship awarded by Bilkent University for undergraduate education, 2002 - 2006.
- Top 0.01% in nationwide University Entrance Exam among 1.5 million candidates, 2002.
- Ranked 1st in the Mediterranean Region in 8th and 9th Turkish Mathematics Olympiad, 2000, 2001.
- Bronze medal in 3rd and 4th Turkish Secondary School Mathematics Olympiad, 1998, 1999.

Activities

- **Program Committee Member:** CIKM 2019
- **Reviewer for Conferences and Journals:** CIKM (2019), KDD (2018), ACM Transactions on Algorithms (2013), Machine Learning (2017), IEEE Transactions on Knowledge and Data Engineering (2013, 2017), Computational Intelligence (2012, 2013), Network Modeling Analysis in Health Informatics and Bioinformatics (2011), Computational and Mathematical Methods in Medicine (2013).
- **Member:** IEEE, ACM.

Affiliations

- Treasurer of Turkish Student Association at RPI, 2011-2012.
- Member of Building Planning Committee, Computer Science, RPI, 2011-2012.
- RPI School of Science Graduate Council, Computer Science Representative, 2010-2011.
- Member of Graduate Recruiting Committee, Computer Science, RPI, 2007-2008.