Muhammad Yaseen Khan

Data Science (Search Experience),
Daraz–Alibaba Group Inc.,
Karachi, Pakistan.

Email: yaseenkhan.yzai@gmail.com
Phone/WhatsApp: (+92) 332 3240104
LinkedIn | Google Scholar | Git

Education

2018 – cont.   Ph.D. (Computer Science), Mohammad Ali Jinnah University, DCS.

2014 – 2015   MS (Software Engineering), Karachi Institute of Economics & Technology, CoC&IS.

2007 – 2010   BS (Software Engineering), University of Karachi, DCS (UBIT).

Employment

01/22 – to date   *Senior Data Scientist – III (Search Experience)* at **Daraz–Alibaba Group Inc.**, Karachi
- Led the data science initiatives for uplifting search experience, for 5 countries and 14 languages/scripts in total.
- Developed (multilingual) NLP solutions for different business problems and challenges.
- Mentored and trained junior resources for ML and NLP skills.
- Collaborated with data analysis, engineering, governance teams in Pakistan, China and Singapore.

02/21 – 01/22   *Principal R&D Engineer* at **Love For Data**, Karachi
- Led the multilingual NLP projects in R&D team.
- Developed FinTech solutions (for an Egyptian micro-loans providing company); and various data profiling and extraction relating projects in different segments of finance, banking that deal with the security factors.
- Collaborated and communicated among different functional teams in overall R&D projects.
- Wrote, estimated and presented the business proposals to clients.

01/17 – 09/22   *Senior Lecturer* at **Mohammad Ali Jinnah University**, Karachi
- Taught undergrad and grad courses.
- Collaborated extensively with in CS faculty members for R&D.
- Organized and worked effectively in the first-ever dedicated lab for language computing.
- Chaired and participated in various boards and committees on student and academic affairs.
- Mentored and supervised different student societies.

01/19 – 12/19   *AI/Software Developer* (part time) at **AskIQ**, Karachi
- Developed FinTech solutions for banking sector (primarily for Standard Chartered Bank).
- Collaborated with different stakeholders for business analysis and project management.

01/16 – 03/17   *Data Scientist* (on contract) at **Cubix**, Karachi
- Developed Python-based Recommendation system for recipe suggestion.
- Collaborated with different functional and development teams for optimized product throughput.

06/13 – 12/16   *Lecturer* at **Karachi Institute of Economics & Technology**, Karachi
- Taught computer science at undergraduate level.
- Collaborated faculty members for Machine Learning and NLP.
- Motivated students for forming the first-ever student society that is ACM-KIET student chapter.
- Managed Final Year Project Committee.

02/12 – 06/13   *Software Engineer – II* at **Mazik Global**, Karachi (in connection with Microsoft Inc., Renton, USA)
- Developed and automated test cases for Microsoft Dynamics (Public Sector).
- Collaborated with Tyler Technologies Inc. and Microsoft Inc. on software testing plans etc.

01/11 – 02/12   *Web+iOS App. Developer* at **Counterpoint**, Karachi (in connection with Qurtaba LLC., Fremont, USA)

Skills

ML/NLP/Data Science	Scikit-Learn · Keras · TensorFlow · NLTK · SpaCy · CoreNLP/Stanza · Gensim · Universal Dependencies · Pandas · Numpy
Scripting/Programming	Python · Java · JavaScript · C# · SQL
Visualization	Matplotlib · Seaborn
Versioning	Git
Typesetting	LaTeX
Languages	English · Urdu · Hindi · Persian/Farsi

Professional Services

Projects

1. **Misspell Correction @ Daraz** - It aims to rectify the misspelt words in search queries for getting an enhanced recall in relevant product retrieval. Approximately 800K misspell-correction pairs are generated (individually and venture-specific; for English, Urdu, Bengali, Nepalese, Burmese and Sinhalese languages in Roman and their native script) that can rectify errors with phonetic and keyboard proximity-based variations. The project brought a positive uplift of 1.14% in CR and 1.77% uplift in search in during AB-testing.

2. **Virtual Title @ Daraz** - For countering the cold start problem for the products that show improper content, low-visibility/low performance, this project aims to erect a parallel/virtual title against such product titles to bring uplift. Various techniques are employed for the creation of virtual titles, which are based on Knowledge Graphs and semantic clustering etc. This project is currently under AB-Testing.

3. **Profanity Filtering and Search Query Optimization @ Daraz** - It aims to remove swear content and profanity related search queries from the search suggestions through a rule-based system. In result, we got 3.89% uplift in CTR.

4. **Kashat Micro-Financing @ Love For Data** - helps customer credit scoring and churn prediction with NLP in/of Arabic and English data; the main features are itemized below:
   - Customer services subscription identification through inbox SMSs.
   - Fake customer identification through fuzzy and exact name matching.
   - Customer defaulter prediction through the SMS exploration.
   - Filtration of SMSs that reflect financial transactions, and the extraction and retrieval of credit and debit amounts exist therein.
   - Segmentation of customer's debit transactions behaviour into payment source (card, wallet, e-payment) and payment targets (grocery, internet, utility, school, etc.)
   - (for the last two features) The average processing time is $<$1000 seconds for 110+ million records.

5. **InstaScan @ Love For Data** - An OCR app for Pakistani CNIC and other identity documents that is capable to register and accommodate CNIC of any size, orientation, camera angle and other visual variations.

6. **Adverse Media @ Love For Data** - Development of automated news aggregating and classification system for checking the adversity and people related to them. It has to fit into a pipeline of criminals and politically exposed persons profiling in financial institutes and data streams.

7. **Standard Chartered Bank's Debit Card Process Automation @ AskiQ** - Helps debit card processing through reading data from different sources and classifying them according to request. The project helped SCB to reduce the manual working of $4-5$ hours to ≈ 3 milliseconds.

8. **e-Grocer Recipe Recommendation @ Cubix** - A fuzzy rule-based search engine/recommender system that process the data (items exists in your pantry) and recommends a whole-week recipe schedule, which needs to be fulfilled accordingly with the number of family members, (and their) calorie intake, allergens, favourites (recipe), likes (ingredients) and dislikes.

9. **ATE Failure Inspection** - Detailed investigation and reporting—to Officials at Tyler Inc., Renton, USA and Microsoft HQ, Redmond, USA—on the automated test execution failures for Microsoft Dynamics R2-PS ERP.

10. **Urdu Writer** - The first Urdu typsetting app in iOS. Downloaded 1.25 million times.

11. **Urdu Poems** - Many iOS applications and alphabet assembling games relate to the children's Urdu education.

Teaching

2017 − 2021   Courses taught to *grad students* in a co-teaching role at MAJU: CS6420 Applied Programming; CS5510 Advanced Research Method, CS6610 Information Retrieval Techniques.
Courses taught to *undergrad students* at MAJU: CS1210 Introduction to Computing; CS1410 Computer Programming; CS1420 Object-Oriented Programming; CS2510 Data Structures and Algorithms; CS3110 Software Engineering; CS3310 Artificial Intelligence; CS5339 Introduction to Data Science.

2013 − 2016   Courses taught to undergrad students at KIET: Programming Fundamentals; Object-Oriented Programming; Data Structures and Algorithms; Data Mining; Software Engineering.

Boards

1. *Member*, Board of Studies, MAJU, Karachi (2017–2019; 2022)

Conference and Workshop Committees

1. *Member*, Final Year Projects External Jury, Suffa University, Karachi (2022)
2. *Member*, Technical committee, Mohammad Ali Jinnah University International Conference on Computing (2021-22)
3. *Member*, Organising committee, International Conference on Data Sciences (2019)
4. *Trainer*, 1-Day Workshop on Data Science with Python at UBIT, Karachi University (2018)
5. *Trainer*, Hands-on Programming with Python at MAJU, Karachi (2018)
6. *Trainer*, Programming with Python, Combat:18 at KIET, Karachi (2018)
7. *Coach*, ICPC 2015 at NUCES-FAST, Lahore (2015)
8. *Faculty Head*, CodeJam at KIET, Karachi (2015)
9. *Faculty Head*, Combat: The Computing Battle 2017 at KIET, Karachi (2014)
10. *Coach*, Startup-Weekend at IBA, Karachi (2012)

University Committees

1. *Member*, Graduate Thesis Evaluation Committee, MAJU, Karachi (2018–present)
2. *Member & Chair*, Examination Paper Moderation Committee at MAJU, Karachi (2018–present; chair, Spring 2020)
3. *Chair*, Committee on Examination Policy for COVID-19 Pandemic at MAJU, Karachi
4. *Member*, QEC Self-Assessment Report for undergrad and grad CS and SE programmes at MAJU, Karachi
5. *Member*, External Course Review Committee at KIET, Karachi (2019)
6. *Member*, Student Academic Advisor/Counselor at MAJU, Karachi (2017–present)
7. *Member*, Final Year Project Evaluation/Exhibition Committee, MAJU, Karachi (2017–21)
8. *Member*, Founding Faculty Sponsor of ACM-MAJU student chapter at MAJU, Karachi (2017–present)
9. *Member*, Admission interviews committee at MAJU, Karachi (2017–present)
10. *Member*, Credit transfer committee at MAJU, Karachi (2018–present)
11. *Member*, Founding Faculty Sponsor of ACM-KIET student chapter at KIET, Karachi (2013–15)
12. *Member*, Final year project committee at KIET, (2013–16)

Publications

Since 2018 (and as of August '22), there are 94 citations with h-index = 6. The Q* statistic of journal articles is Q1=3, Q2=1, Q3=4 and Q4=1. Published items are sorted chronologically in reverse order. Conference papers that are later converted into book chapters or journal articles are counted once.

Refereed Journal Articles

1. Muhammad Yaseen Khan, Tafseer Ahmed, Shaukat Wasi, and Muhammad Shoaib Siddqui. "Enhancing Sarcasm and Sentiment Analysis with Cognitive Relationship: A Context-Aware Approach for Urdu—a Resource Poor Language" *Computational Intelligence and Neuroscience* (2022). (Q1; SCI-IF 3.6) [PDF. In production]

2. Tafseer Ahemd, Muhammad Suffian, Muhammad Yaseen Khan, and Alessandro Bogliolo. "Discovering Lexical Similarity Using Articulatory Feature-Based Phonetic Edit Distance." *IEEE Access* (2021). (Q1; SCI-IF 3.36) [PDF]

3. Muhammad Yaseen Khan, Abdul Qayoom, Muhammad Suffian Nizami, Muhammad Shoaib Siddiqui, Shaukat Wasi, and Syed Muhammad Khaliq-ur-Rahman Raazi. "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques." *Complexity* 2021 (2021). (Q1; SCIE-IF 2.83) [PDF]

4. Muhammad Suffian Nizami, Tafseer Ahmed, and Muhammad Yaseen Khan. "Hindustani or Hindi vs. Urdu: A Computational Approach for the Exploration of Similarities Under Phonetic Aspects." *International Journal of Advanced Computer Science and Applications*, 11(11), 2020. (Q3; ESCI) [PDF]

5. Muhammad Yaseen Khan, Muhammad Adil Rao, Shaukat Wasi, Twaha Ahmed Minai, and Syed Muhammad Khaliq-ur-Rahman Raazi. "Edit distance-based search approach for retrieving element-wise prosody/rhymes in Hindi-Urdu poetry." *Indian Journal of Science and Technology* 13(39), 2020: pp. 4189–4201. (MJL) [PDF]

6. Muhammad Yaseen Khan and Khurum Nazir Junejo. "Exerting 2D-Space of Sentiment Lexicons with Machine Learning Techniques: A Hybrid Approach for Sentiment Analysis." *International Journal of Advanced Computer Science and Applications* 11(6), 2020: pp. 599–608. (Q3; ESCI) [PDF]

7. Waqar Muhmamd, Maria Mushtaq, Khurum Nazir Junejo, and Muhammad Yaseen Khan. "Sentiment Analysis of Product Reviews in the Absence of Labelled Data using Supervised Learning Approaches." *Malysian Journal of Computer Science* 32(2), 2020: pp. 118–132. (Q3; SCI-IF 0.7) [PDF]

8. Sham Muhammad Emad, Kamran Ahsan, Muhammad Yaseen Khan, Shoaib Ahmed, Rafiullah Khan, Kashif Rizwan, Farhan Shafiq. "Depression Detection using Naïve Bayes: Text-based approach." *Interciencia* 44(11), 2020: pp. 174–182. (Q2; SCIE-IF 0.26)

9. Kanwal Iqbal, Muhammad Yaseen Khan, Shaukat Wasi, Shumaila Mahboob, and Tafseer Ahmed. "On Extraction of Event Information from Social Text Stream: An Unpretentious NLP Solution." *International Journal of Computer Science and Network Security* 19(9), 2019: pp. 121–131. (ESCI) [PDF]

10. Aisha Nazir, Muhammad Yaseen Khan, Tafseer Ahmed, Syed Imran Jami, and Shaukat Wasi. "A Novel Approach For Ontology-Driven Information Retrieving Chatbot For Fashion Brands." *International Journal of Advanced Computer Science and Applications* 10(9), 2019: pp. 546–552. (Q3; ESCI) [PDF]

11. Muhammad Suffian, Muhammad Yaseen Khan, and Shaukat Wasi. "Developing Disease Classification System based on Keyword Extraction and Supervised Learning." *International Journal of Advanced Computer Science and Applications* 9(9), 2018: pp. 599–605. (Q4; ESCI) [PDF]

Refereed Book Chapters

12. Muhammad Yaseen Khan and Tafseer Ahmed. "Pseudo transfer learning by exploiting monolingual corpus: An experiment on Roman Urdu transliteration." *International Conference on Intelligent Technologies and Applications*. Springer, Singapore, 2019. [PDF]

13. Muhammad Suffian Nizami, Muhammad Yaseen Khan, and Tafseer Ahmed. "Towards a generic approach for pos-tagwise lexical similarity of languages." *International Conference on Intelligent Technologies and Applications*. Springer, Singapore, 2019. [PDF]

Refereed Conference Publications

14. Abdul Basit, Muahmmad Yaseen Khan, Syed Sarmad Ali, Muhammad Suffian, and Sumra Khan. "Gender Classification Using Smartphone Sensors and Machine Learning Techniques" In *2022 17th International Conference on Emerging Technologies (ICET)*, GIKI, Pakistan.

15. Muhammad Suffian, Muahmmad Yaseen Khan, and Alessandro Bogliolo. "Towards Human Cognition Level-based Experiment Design for Counterfactual Explanations (XAI)" In 2022 *2nd Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, Karachi, Pakistan.

16. Muahmmad Yaseen Khan and Muhammad Umer Tariq Butt. "Classic Fairy Tales / Children Stories of the World: Corpus Organization and Text Analytics." In 2021 *7th International Conference of Linguistic Association of Pakistan (ICLAP)*, Sargodha, Pakistan.

17. Samiullah Shaikh, Muhammad Yaseen Khan, and Muhammad Suffian Nizami. "Using Patient Descriptions of Top-20 Diseases in Text Classification for Evidence-based Medicine." In 2021 *1st Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, pp. 1–8. IEEE, 2021. Karachi, Pakistan. [PDF]

18. Kashif Badr, Kanwal Iqbal, Shumaila Mahboob, and Muhammad Yaseen Khan. "Processing Fire Detection, Face Recognition, and Head Counting for an IoT-based Application in Emergency Exit." In 2021 *1st MAJU International Conference on Computing (MAJICC)*, pp. 1–6. IEEE, 2021. Karachi, Pakistan. [PDF]

19. Sohail, Muhammad Yaseen Khan, Nisar Ahmed Siddiqui, and Kiran. "Sign Language Translation in Urdu/Hindi Through Microsoft Kinect." In 2020 *3rd Pak-Turk International Conference on Emerging Technologies in the field of Sciences and Engineering (ETSE)*, pp. 1–7. IOP Conference Series: Material Sciences and Engineering, 2020. GIK Topi, Pakistan. [PDF]

20. Muhammad Yaseen Khan and Muhammad Suffian Nizami. "Urdu Sentiment Corpus (v1.0): Linguistic Exploration and Visualization of Labeled Dataset for Urdu Sentiment Analysis." In 2020 *2nd International Conference on Information Science & Communication Technology (ICISCT)*, pp. 1–15. IEEE, 2020. Karachi, Pakistan. [PDF]

21. Mohammad Ayub Latif, Muhmmad Yaseen Khan, and Kashif Bashir. "Practices for Achieving Accuracy in Software Costing and Estimation." In *2018 2nd International Conference on Computing & Information Sciences (ICCIS)*, pp. 83–95. IEEE, 2018. Karachi, Pakistan. [PDF]

22. Shumaila Mahboob, Kanwal Iqbal, Hira Farman, Muhammad Yaseen Khan, and Shaukat Wasi. "Drive pattern analyzer using smartphones." In *2017 13th International Conference on Emerging Technologies (ICET)*, pp. 1–5. IEEE, 2017. Islamabad, Pakistan. [PDF]

23. Muhammad Yaseen Khan, Shah Muhammad Emaduddin, and Khurum Nazir Junejo. "Harnessing English sentiment lexicons for polarity detection in urdu tweets: A baseline approach." In *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, pp. 242–249. IEEE, 2017. San Diago, USA. [PDF]

Graduated Students

1. Muhammad Suffian Nizami (MS Thesis - CS); "*Formulating Patient Descriptions to Support Evidence-Based Medicine*". Spring 2018. Co-supervised with Dr S. Wasi.
2. Alvina Naveed (MS Thesis - SE); "*Exploiting Bug Narrations for Predictive Analysis on Bug Severity*". Fall 2020.
3. Abdul Basit (MS Thesis - CS); "*Human motion tracking for gait analysis through smartphones and machine learning approaches*". Fall 2020.
4. Abdul Qayoom Bhatti (MS Thesis - CS); "*Supervised Learning-based Good Dictionary Examples Recommendation System*". Fall 2020.
5. Mirza Muhammad Muneeb Baig (MS Thesis - CS); "*Sarcasm detection in Urdu Text*". Fall 2020.

Awards and Honours

1. Awardee of Dean's List 2014 and 2015 during MS.SE, KIET

References

1. Dr. Khurum Nazir Junejo, former Head of R&D at Love for Data.
2. Dr. Tafseer Ahmed Khan, former Professor and Head Center for Language Computing, MAJU, Karachi.
3. Dr. Shaukat Wasi, Professor and Dean Faculty of Computing, MAJU, Karachi.
4. Dr. Asim Imdad Wagan, Professor and Dean Faculty of Engineering, MAJU, Karachi.
5. Dr. Muhammad Khalid Khan, Professor and Director CoCIS, PAF-KIET, Karachi.