

Muhammad Yaseen Khan

Curriculum Vitæ

Research & Development
Love For Data
Pakistan

Tel: +92-332-3240104
Email: yaseenkhan.yzai@gmail.com
Skype: muhammad_yaseen_khan



Education

PhD (CS)	Mohammad Ali Jinnah University	2018–cont'd.	Advisors: Dr. Tafseer Ahmed and Dr. Shaukat Wasi
MS (SE)	Karachi Institute of Economics & Technology	2014–2015	Advisor: Dr. Khurram N. Junejo
BS (SE)	University of Karachi	2007–2010	Advisors: Dr. Humera Tariq and Dr. Tahseen A. Jilani

Areas of Interest

Computational Linguistics · Natural Language Processing (NLP) · Information Extraction · Information Retrieval · Lexical Semantics · Emotions and Sentiment Analysis · Information Visualization · Psycholinguistics · Computational Social Science.

Positions Held

Current

Feb. 2021–Present *Principal R&D Engineer* Love For Data, Karachi, Pakistan

Previous

Jan. 2017–Feb. 2021	<i>Senior Lecturer</i>	DCS, FoC, Muhammad Ali Jinnah University, Karachi, Pakistan
Jan. 2019–Dec. 2019	<i>AI/Software Developer</i>	AskIQ, Karachi, Pakistan
Jan. 2016–Mar. 2017	<i>Data Scientist</i>	Cubix Co., Karachi, Pakistan
Jun. 2013–Dec. 2016	<i>Lecturer</i>	CoCIS, Karachi Institute of Economics & Technology, Khi, Pakistan
Feb. 2012–Jun. 2013	<i>Software Engineer II</i>	Mazik Global Technologies, Khi, Pakistan / Microsoft Inc., Renton, USA
Jan. 2011–Feb. 2012	<i>Web+iOS App. Developer</i>	Counterpoint, Karachi, Pakistan / Qurtaba LLC., Fremont, USA
Jul. 2010–Dec. 2010	<i>Web Developer</i>	Revolutionary Technologies, Karachi, Pakistan
Jan. 2010–Jun. 2010	<i>Web Developer (Intern)</i>	SprintTechSolutions, Karachi, Pakistan

Project Highlights

Industrial Projects

1. *Kashat Micro-Financing* - 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.
2. *InstaScan* - 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.
3. *Adverse Media* - 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.
4. *Standard Chartered Bank's Debit Card Process Automation* - 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.
5. *e-Grocer Recipe Recommendation* - 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.

6. *Qoot Restaurant-Rider Pairing* - Algorithmic designing of Saudi-Arabian delivery app that does optimal match-making between two restaurants and targets for a single delivery person given time and distance constraints.
7. *ATE Failure Inspection* - Detailed investigation and reporting on the automated test execution failures for Microsoft Dynamics R2-PS ERP. The reporting hierarchy includes Tyler officials from Renton, USA and Microsoft HQ at Redmond, USA.
8. *Urdu Writer* - The first Urdu type app in iOS. Downloaded 1.25 million times.
9. *Urdu Poems* - Many iOS applications and alphabet assembling games relating to the children Urdu education.

Academic/Research Projects

1. Sarcasm and its co-relation with the sentiments in Urdu Text.
2. Automated classification of software bugs for their severity and priority.
3. Walk-Pattern analysis for gait diseases and gender classification.
4. Gender-based literature exploration in Urdu micro-fiction and short stories.
5. Affective computing in Urdu text.
6. Prosodic-term retrieval in Urdu/Hindi poetry.
7. Roman Urdu transliteration with character-based LSTM.
8. Automated GDEx elicitation with machine learning and word embedding-based deep neural network.
9. Lexical comparison of Hindi and Urdu.
10. Phonetic Edit-Distance and its application for checking language commonality between Arabic, Hebrew, Sanskrit, Marathi, Persian, and Urdu.
11. For more information, please see [publications](#).

Teaching

- 2017–present Courses taught to *grad students* 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.

Conference and Workshop Committees

1. *Member*, Technical committee, 1st Mohammad Ali Jinnah University Intl. Conf. on Computing, MAJU, Karachi (2021)
2. *Member*, Organising committee, 1st Intl. Conf. on Data Sciences, MAJU, Karachi (2019)
3. *Trainer*, 1-Day Workshop on Data Science with Python, UBIT, Karachi University (2018)
4. *Trainer*, Hands-on Programming with Python, MAJU, Karachi (2018)
5. *Trainer*, Programming with Python, Combat:18, KIET, Karachi (2018)
6. *Coach*, ICPC 2015 @ NUCES-FAST, Lahore (2015)
7. *Faculty head*, CodeJam, PAF-KIET, Karachi (2015)
8. *Faculty head*, Combat: The Computing Battle 2017, KIET, Karachi (2014)
9. *Coach*, Startup-Weekend, IBA, Karachi (2012)

University Committees

1. *Member & Chair*, Examination paper moderation committee, MAJU, Karachi (2018–present; *chair*, Spring 2020)
2. *Chair*, Committee on examination policy for COVID-19 pandemic, MAJU, Karachi
3. *Member*, QEC Self-Assessment Report for undergrad and grad CS and SE programmes, MAJU, Karachi
4. *Member*, External course review committee, KIET, Karachi (2019)
5. *Member*, Graduate thesis evaluation committee, MAJU, Karachi (2018–present)

6. *Member*, Academic advisor, MAJU, Karachi (2017–present)
7. *Member*, Final year project evaluation/exhibition committee, MAJU, Karachi (2017–present)
8. *Member*, Founding faculty sponsor of ACM-MAJU student chapter, MAJU, Karachi (2017–present)
9. *Member*, Admission interviews committee, MAJU, Karachi (2017–present)
10. *Member*, Credit transfer committee, MAJU, Karachi (2018–present)
11. *Member*, Faculty sponsor of ACM-KIET student chapter, KIET, Karachi (2013–2015)
12. *Member*, Final year project committee, KIET, (2013–2016)

Boards

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

Publications

Items are sorted chronologically in reverse order. Conference papers that later converted into book chapters or journal articles are counted once.

Refereed Journal Articles

- J1. “Discovering Lexical Similarity Through Articulatory Feature-based Phonetic Edit Distance”. T. Ahmed, M. S. Nizami, M. Y. Khan, and A. Bogliolo; *IEEE Access*, volume 2021, 2021. Q1 SCIE IF: 3.36 *accepted*.
- J2. “Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning and Word Embedding-based Deep Learning Techniques”. M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi, and S. M. K. R. Raazi; *Complexity*, volume 2021, 2021. Q1 SCIE IF: 2.83
- J3. “Hindustani or Hindi vs. Urdu: A Computational Approach for the Exploration of Similarities Under Phonetic Aspects”. M. S. Nizami, T. Ahmed, and M. Y. Khan; *International Journal of Advanced Computer Science and Applications*, volume 11, number 11, pages 749–755, 2020. Q3 ESCI
- J4. “Edit Distance-based Search Approach for Retrieving Element-wise Prosody in Hindi/Urdu Poetry”. M. Y. Khan, M. A. Rao, S. Wasi, T. A. Minai; *Indian Journal of Science and Technology*, volume 13, number 39, pages 4189–4201, 2020. MJL indexed
- J5. “Exerting 2D-Space of Sentiment Lexicons with Machine Learning Techniques: A Hybrid Approach for Sentiment Analysis”. M. Y. Khan and K. N. Junejo; *International Journal of Advanced Computer Science and Applications*, volume 11, number 6, pages 599–608, 2020. Q3 ESCI
- J6. “Sentiment Analysis of Product Reviews in the Absence of Labelled Data using Supervised Learning Approaches”. W. Muhammad, M. Mushtaq, K. N. Junejo, and M. Y. Khan; *Malaysian Journal of Computer Science*, volume 32, number 2, pages 118–132, 2020. Q3 SCI IF: 0.7
- J7. “Depression Detection using Naïve Bayes: Text based approach”. S. M. Emad, K. Ahsan, M. Y. Khan, S. Ahmed, R. Ullah, K. Rizwan, and F. Shafiq; *Interciencia*, volume 44, number 11, pages 174–182, 2020. Q2 SCIE IF: 0.26
- J8. “On Extraction of Event Information from Social Text Stream: An Unpretentious NLP Solution”. K. Iqbal, M. Y. Khan, S. Wasi, S. Mahboob, and T. Ahmed; *International Journal of Computer Science and Network Security*, volume 19, number 9, pages 121–131, 2019. ESCI
- J9. “A Novel Approach For Ontology-Driven Information Retrieving Chatbot For Fashion Brands”. A. Nazir, M. Y. Khan, T. Ahmed, S. I. Jami, and S. Wasi; *International Journal of Advanced Computer Science and Applications*, volume 10, number 9, pages 546–552, 2019. Q3 ESCI
- J10. “Developing Disease Classification System based on Keyword Extraction and Supervised Learning”. M. Suffian, M. Y. Khan, and S. Wasi; *International Journal of Advanced Computer Science and Applications*, volume 9, number 9, pages 599–605, 2018. ESCI

Refereed Book Chapters

- B1. “Pseudo Transfer Learning by Exploiting Monolingual Corpus: An Experiment on Roman Urdu Transliteration”. M. Y. Khan and T. Ahmed; in: Bajwa I., Sibaliya T., Jawawi D. (eds) *Intelligent Technologies and Applications*. INTAP 2019. *Communications in Computer and Information Science*, vol 1198. Springer, Singapore.

- B2. “Towards a Generic Approach for PoS Tag-wise Lexical Similarity of Languages”. M. S. Nizami, M. Y. Khan and T. Ahmed; in Bajwa I., Sibalija T., Jawawi D. (eds) Intelligent Technologies and Applications. INTAP 2019. *Communications in Computer and Information Science*, vol 1198. Springer, Singapore.

Refereed Conference Publications

- C1. “Classic Fairy Tales / Children Stories of the World: Corpus Organization and Text Analytics”. M. Y. Khan and M. U. T. Butt; in *Proceedings of the 7th Intl. Conf. of Linguistic Association of Pakistan (ICLAP '21)*, Sargodha, Pakistan. Nov. 2021.
- C2. “Using Patient Descriptions of Top-20 Diseases in Text Classification for Evidence-based Medicine”. S. Shaikh, M. Y. Khan, and M. S. Nizami; in *Proceedings of the 1st MAJU Intl. Conf. on Computing (MAJICC '21)*, Karachi, Pakistan. July. 2021.
- C3. “Processing Fire Detection, Face Recognition, and Head Counting for an IoT-based Application in Emergency Exit”. K. Badr, K. Iqbal, S. Mahboob, and M. Y. Khan; in *Proceedings of the 1st MAJU Intl. Conf. on Computing (MAJICC '21)*, Karachi, Pakistan. July. 2021.
- C4. “Sign Language Translation in Urdu/Hindi Through Microsoft Kinect”. Sohail, M. Y. Khan, N. A. Siddiqui, and Kiran; in *Proceedings of the 3rd Pak-Turk Conference On Emerging Technologies*, GIKI, Pakistan. June. 2020.
- C5. “Urdu Sentiment Corpus (v1.0): Linguistic Exploration and Visualization of Labeled Dataset for Urdu Sentiment Analysis”. M. Y. Khan and M. S. Nizami; in *Proceedings of the 2nd IEEE Intl. Conf. on Information Science & Communication Technology*, Karachi, Pakistan. Feb. 2020.
- C6. “Practices for Achieving Accuracy in Software Costing and Estimation”. M. A. Latif, M. Y. Khan, and K. Bashir; in *Proceedings of the 2nd Intl. Conf. on Computing & Information Sciences*, KIET, Karachi, Pakistan. Nov. 2018.
- C7. “Drive pattern analyzer using smartphones”. S. Mahboob, K. Iqbal, H. Farman, M. Y. Khan, and S. Wasi; in *Proceedings of the 13th IEEE Intl. Conf. on Emerging Technologies*, CUST, Islamabad, Pakistan. IEEE, pages 1-5, 2017.
- C8. “Harnessing English Sentiment Lexicons for Polarity Detection in Urdu Tweets: A Baseline Approach”. M. Y. Khan, S. M. Emaduddin, and K. N. Junejo; in *Proceedings of the 11th IEEE Intl. Conf. on Semantic Computing*, San Diego, USA. IEEE, pages. 242-24, 2017.

Unpublished Work

- U1. “Developing Sentiment Lexicon and Labelled Corpus for Urdu Sentiment Analysis”. M. Y. Khan, R. Zaheer; *experimentation completed; in draft*.
- U2. “Developing Emotion Lexicon and Labelled Corpus for Urdu Emotion Detection”. M. Y. Khan, T. Khalid; *experimentation completed; in draft*.
- U3. “Urdu Named-Entity Recognition System”. T. Ahmed, M. Y. Khan, and S. Nizami; *experimentation completed; in draft*.
- U4. “LSTM-based Urdu Word Segmentation”. M. Y. Khan and T. Ahmed; *experimentation completed; in draft*.
- U5. “Depressive Sentiment Analysis in Microblogs: A Machine Learning Approach”. S. M. Emaduddin, K. Ahsan, M. Y. Khan, and S. Ahmed; *under review*.
- U6. “Efficient Recipe Recommendation System Using Signature-Trees”. A. Sohani, M. Y. Khan, and R. Ullah; *accepted with revision*.

Conference and Workshop Attendee

- 7th Intl. Conf. of Linguistics of Pakistan (ICLAP), Univesity of Sargodha, Pakistan (2021)
- 1st Mohammad Ali Jinnah Univesity Conf. on Computing (MAJICC), Karachi, Pakistan (2021)
- 7th Intl. Conf. on Language and Technology (CLT), CLE-UET, Lahore (2020)
- 2nd Intl. Conf. on Information Science & Communication Technology (ICISCT), UoK, Karachi, Pakistan (2020)
- 2nd Intl. Conf. on Intelligent Technologies and Applications (INTAP), IUB, Bhahawalpur (2019)
- 1st Intl. Conf. on Data Science, MAJU, Karachi, Pakistan (2019)
- 2nd Intl. Conf. on Computing and Information Sciences, KIET, Karachi, Pakistan (2018)
- HEC-NCEAC 3-Day Workshop on Data Science, Islamabad (2017)

Graduated Students

- Muhammad Suffian Nizami (MS Thesis - CS); “Formulating Patient Descriptions to Support Evidence Based Medicine”. Spring 2018. Co-supervised with Dr S. Wasi.
- Alvina Naveed (MS Thesis - SE); “Exploiting Bug Narrations for Predictive Analysis on Bug Severity”. Fall 2020.
- Abdul Basit (MS Thesis - CS); “Human motion tracking for gait analysis through smart phones 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

Skills

Computing Programming (Python, Java, JavaScript, C#, MySQL) · Typesetting with \LaTeX
 Languages English · Urdu · Hindi · Persian/Farsi

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](#), CIO AskIQ; Professor and Dean Faculty of Engineering, MAJU, Karachi.
5. [Dr. Muhammad Khalid Khan](#), Professor and Director CoCIS, PAF-KIET, Karachi.