

SABIR ISMAIL

Address: Stony Brook, NY, 11794 ◊ Phone: +718-213-5127 ◊ Email: saismail@cs.stonybrook.edu

🔗 <https://github.com/sabir4063> ◊ **in** <https://www.linkedin.com/in/sabir-ismail-83113b20>

EDUCATION

Masters of Engineering in Computer Science

Aug. 2017 – May. 2019 (Expected)

Stony Brook University, Stony Brook, NY

Relevant Courses: Data Science Fundamental, Big Data Analysis, Natural Language Processing, Theory of Database System, Artificial Intelligence, Discrete Mathematics, Computational Biology

Masters of Engineering in Computer Science and Engineering

Jan. 2011 – Aug. 2013

Shahjalal University of Science and Technology Sylhet, Bangladesh

Bachelors of Engineering in Computer Science and Engineering

Jun. 2005 – Dec. 2009

Shahjalal University of Science and Technology Sylhet, Bangladesh

EXPERIENCE

Research Assistant

Jan. 2017 - Jan. 2018

Stony Brook University

Stony Brook, NY

- **Natural Language Processing and Text Mining:** I analyze the posts from the online forums, used by the screen reader user community and identified the issues and problems regarding the use of the different software.

Assistant Professor

Aug. 2013 - Aug. 2017

Shahjalal University of Science and Technology (SUST)

Sylhet, Bangladesh

- **Course Instructor:** Machine Learning, Bio-informatics Computing, Algorithm Design and Analysis, Data Structures, Structured Programming.
- **SUST Natural Language Processing (NLP) Research Group:** I conduct research along with students on Bengali language processing on various topics including stemming, part-of-speech tagging, anaphora resolution, stylometric analysis, automated question answering system. As an outcome of these researches, I have 16 conference and 4 journal papers.
- **Admission Test Automation:** I have worked in the automation of the undergraduate admission test processing system in SUST, which contain the implementation of the full admission process, start from the submitting the application, result processing and end up with the admitting of the students.
- **Center for research, testing and consultancy (CRTC) SUST:** As a member of CRTC, I lead a team to design and develop an online platform (www.konnect.edu.bd) for the school students to engage in different educational and creative activity.
- **Medical College Admission Test Result Processing System:** I have to lead a team to implement the result processing system of centralized admission test for the all the medical college in Bangladesh.

Assistant Professor

Oct. 2010 - Aug. 2013

Metropolitan University

Sylhet, Bangladesh

- **Course Instructor:** Machine Learning, Object Oriented Programming, Algorithm Design and Analysis, Data Structures, Structured Programming, Discrete Mathematics.

Software Engineer

Mar. 2010 - Oct. 2010

Structured Data System Limited

Dhaka, Bangladesh

- **AfriGIS Blackberry Navigator:** Implemented real-time navigation system (also simulation) for Blackberry mobile device using Java blackberry native API.

SKILLS

Programming Language

C, C#, JAVA, Python, Matlab

Database and Libraries

MySQL, Spark, Scikit-learn, Keras, Pandas, NumPy, NLTK, SciPy

PROJECTS

Economic Trends Assessment Using the Night Light Image (Python, Spark):

- Extracted the country region from the night light satellite images from 1992 to 2017 and perform the calibration on the images of the different years.

- For the target country, collect different socio-economic indicator and analyze the correlation between increase of the night light image intensity.

Uncovering the Accessibility Problems in the Systems And Software (Python):

- Collect posts from two screen reader user online forums and perform topic modeling to identify the issues related with the using different software.
- Extracted most used software by the blind people and their problem when using screen reader in data driven approach.

Predicting the Likelihood of a Book (Python)

- Generate word embedding representation of a book and analyze it's association with different set of concepts words (gender, time, money etc).
- Implement a predictive model based on the association and predict the likelihood of the book.

Classification and Entity Extraction from Public Notices (Python)

- Perform entity recognition on 100k public notices from 7 different states and extracted different entities form the notices, including who is the subject and who is the attorney etc.
- Implement classification model to categorized the notices into different category such as auction notice, article of organization, summons, taxes etc.
- Preform the correlation analysis with different socio-economic factor such as population, employment, poverty, income etc and uncover underlying social trends.

NYC Taxi Fare Prediction (Python)

- A Kaggle contest to predict NYC taxi fare and I ranked 159 out of 1230 participant.
- Perform exploratory data analysis and build regression and gradient boosting models with different sets of features.

PUBLICATIONS

- Shamim Ehsan, Sadia Tasnim Swarna, and **Sabir Ismail**. "*A Crowdsourcing-Based Approach for Preparing Bangla POS Tagged Corpus*." In Emerging Technologies in Data Mining and Information Security, pp. 463-473. Springer, Singapore, 2019.
- Md Montaser Hamid, Tanvir Alam, **Sabir Ismail**, and Md Forhad Rabbi. "*Bangla Interrogative Sentence Identification from Transliterated Bangla Sentences*." In 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), pp. 1-6. IEEE, 2018.
- Shuvanon Razik, Evan Hossain, **Sabir Ismail**, and Md Saiful Islam. "*Sust-bhnd: A database of bangla handwritten numerals*." In Imaging, Vision & Pattern Recognition (icIVPR), 2017 IEEE International Conference on, pp. 1-6. IEEE, 2017.
- Tapashee Tabassum Urmi, Jasmine Jahan Jammy, and **Sabir Ismail**. "*A corpus based unsupervised Bangla word stemming using N-gram language model*." In Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on, pp. 824-828. IEEE, 2016.
- Prapti Das, Rishmita Tasmim, and **Sabir Ismail**. "*An experimental study of stylometry in Bangla literature*." In Electrical Information and Communication Technology (EICT), 2015 2nd International Conference on, pp. 575-580. IEEE, 2015.
- **Sabir Ismail**, and M. Shahidur Rahman. "*Bangla word clustering based on N-gram language model*." In Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on, pp. 1-5. IEEE, 2014.

COMPETITIVE PROGRAMMING

- I regularly participated in the competitive programming and I have solved more then 600 problems on online sites during my undergraduate study.
- I also participated in the ACM ICPC Dhaka regional contest and National Collegiate Programming Contest in 2007 and 2008.

HONORS & AWARDS

- 9'th place in National Collegiate Programming Contest 2007.
- Educational Board Scholarship in 8th, 10th, and 12th-grade class.
- Departmental scholarship in every semester for obtaining highest GPA.