KILOL GUPTA

New York NY 10025 | https://www.linkedin.com/in/kilol-gupta/| kilol.gupta@columbia.edu | https://github.com/kilolgupta| (929) 310-7782

EDUCATION

Columbia University, New York - *Masters, Computer Science (GPA: 3.9)*   Sep'17 – Dec'18 (Estimated)
- Coursework: Natural Language Processing, Machine Learning, Deep Learning, Adv. Machine Learning - Personalization, Computational Models of Social Meaning
- Graduate TA Fellowship awardee (awarded to 10 out of 250 students), WiCS GHC'18 Scholarship Recepient
- TA for Cloud Computing and Big Data, Advanced Software Engineering, Introduction to Databases

Delhi Technological University, India - *Bachelors, Computer Science (GPA: 4.0)*   Jul'11 – April'15
- Coursework: Pattern Recognition; Artificial Intelligence; Operating Systems; Databases; Algorithms

EXPERIENCE

IBM T.J. Watson Research Center, Incoming ML-NLP Research Intern   Sep'18 – Dec'18
Electronic Arts, Data (and Software) Engineer Intern   May'18 – Aug'18
- Designed scalable Player Insight Network ETL pipelines to support dynamic priority-based processing of games' data using JAVA, Apache Hive, Oozie, Spark
- Performed data augmentation and analysis for the automation of data ingestion pipeline of a speech synthesis project using Docker and Python

Columbia University, Graduate Research Assistant   Sep'17 – Dec'17
- Implemented coreference resolution for true-case and caseless text and compared performance with CoNLL dataset. Used JAVA, Python, Stanford coreNLP API: (Github)
- Formalized a modified Levenshtein algorithm using NLP techniques to find a match between variations of the same location/organization/person entity. For e.g. Barack Obama and Barack Hussein Obama: (Github)

Royal Bank of Scotland, Software Engineer   Jul'15 – Jul'17
- Implemented Apache Hive UDFs and UDAFs for an Apache Spark powered data warehouse
- Proposed optimizations using Apache Parquet's filter, predicate pushdown. CSV file ingestion into Parque

PROJECTS

- **Deep Attention based Abusive Tweets classification**: Implemented contextual attention based BiLSTM model augmented with LIWC features and user embeddings; Submitted paper in EMNLP 2018, 2nd workshop in Abusive language (Github)
- **Cross-domain Claim Identification in argumentation text using Deep Learning:** Implemented a hybrid attention-based model to identify claim from argumentative text across 6 datasets. (Github)
- **Hybrid Variational Autoencoder for Collaborative Filtering:** Implemented a hybrid VAE on MovieLens 20M and Spotify million playlist dataset. Visualized user and movie embeddings', clustering using k-means and t-SNE (Github)
- **Contextual-Bandit based personalized news article recommendation system:** Implemented LinUCB algorithm for news recommendation (Github)
- **Transition based dependency parsing using feed-forward neural network:** Trained a feed-forward neural network using Dynet Python library and achieved a test accuracy of 84% (Github)
- **Facial Emotion Recognition using Entropy based Encoding:** Applied image encoding schemes in feature extraction on FEEDTUM image database, classification using SVMs & binary clustering using Fuzzy c-means (PDF)
- **Comparative analysis of Neo4j and MySQL via modelling a movie identification system:** Submitted a research paper in ISEC 2017, ACM India. Awarded highest marks in the department (PDF)

PUBLICATIONS

"A Hybrid Variational Autoencoder for Collaborative Filtering", Deep Learning Day at KDD 2018- (KDD Website)

LANGUAGES AND TECHNOLOGIES

- **Languages:** JAVA, Python, SQL, NodeJS   **Familiar:** C++, C#, C, Scala, JavaScript, Latex
- **Deep Learning Frameworks:** Keras, TensorFlow, Theano, Dynet
- **IDE**: Intellij IDEA, Eclipse, Microsoft Visual Studio
- **Big Data:** Neo4j, MongoDB, Apache Spark**,** Zookeeper, Hive, Parquet, HDFS, Yarn, Oozie, ORC file format
- **Google Cloud, AWS**: Amazon ML, Lex, Elastic Beanstalk, Lambda Functions, S3, SNS, SQS, API Gateways, Orchestrators