Himanshu Taneja

(979) 985 8727, 027himanshu@gmail.com
401 Stasney St Apt 304, College Station, Texas 77840

OBJECTIVE

Seeking summer internship opportunities in the fields of data science, machine learning, and natural language processing

EDUCATION

Texas A&M University, College Station, Texas *May 2018*
Masters of Science in Electrical Engineering, GPA: 4.0
Courses: Linear Algebra, Statistical Theory, Regression Analysis, Natural Language Processing, Structural Bioinformatics

USICT, Guru Gobind Singh Indraprastha University, Delhi, India *May 2016*
Bachelor of Technology in Electronics and Communications Engineering, GPA: 74.85/100 (3rd in class)

SKILLS

Programming Languages
Most experienced with C++, Java, Python, R, Matlab
Some experience in C, Bash, Awk, SQL

Tools and Libraries
Git, Regular Expressions, Apache Spark, Microsoft Excel, Relational Databases, NLTK, scikit-learn

PROJECTS

Recommendation System for News Stories (Thesis) *Ongoing*
- Gathering and analyzing article structure of news stories from digital platforms such as Reddit, CNN & FoxNews
- Identifying the attributes that differentiate front page news from other stories; to design a model that can suggest top news stories to the editors

Part-of-Speech Tagger *Spring '17*
- Implemented a Hidden-Markov model in Python to tag words in a text corpus as Noun, Verb, Preposition etc.
- After initialization the model can improve upon itself by learning from the already tagged words.

Molecular Conformation Optimization *Spring '17*
- Designed a Monte-Carlo sampling method in Python to optimize the placement of atoms in 3-dimensional structure of a molecule.
- The application is scalable to any number of atoms requiring only 3 bytes of storage per atom and has a constant-time complexity per iteration

Sequence Aligner *Spring '17*
- Implemented the Needleman-Wunsch algorithm in Python to align Protein and DNA sequences
- The application can scan through a database of sequences and return the best match possible for a target sequence

Classification of Stacking Fault Energy of Alloys *Fall '16*
- Analyzed the effect of chemical composition of Steel alloys on their Stacking Fault Energy (SFE)
- Identified most significant elements affecting SFE of alloys using T-test and Principal Component Analysis
- Trained and benchmarked the classification algorithms (Linear Discriminant Analysis, K-Nearest Neighbors, and Support Vector Machines) on the dataset

Automatic Text Classification and Summarization *Summer '15*
- Developed a text classifier using machine learning algorithms (Naive Bayes, K-Nearest Neighbors, and Support Vector Machines)
- Designed a bootstrapping approach to generate summary of news article using their headlines
- Techniques used: stop-words removal, stemming, term frequency-inverse document frequency, additive smoothing

Image Enigma: Encrypt Digital Images *Spring '15*
- Implemented the Enigma Machine (a polyalphabetic cipher) in Python and engineered it to encrypt digital images

Person of Interest *Spring '15*
- Analyzed Enron dataset in Python using machine learning algorithms to identify persons of interest in the Enron Scandal

VocabList: A cross-platform application *Spring '15*
- Developed a cross-platform GUI application using Kivy Framework in Python to maintain a database of words for improving vocabulary

ACTIVITIES

TechSpace (Technical Club at USICT) *Jan 2014 – May 2016*
- Organized InfoXpression (Annual technical fest) and monthly LAN Gaming Contests
- Presented seminar on "Python Programming Language in Data Science"