**SANMAN YADAV**

(315)-378-2220 | [smyadav@syr.edu](mailto:smyadav@syr.edu) | [www.linkedin.com/in/sanman-yadav](http://www.linkedin.com/in/sanman-yadav)

**EDUCATION:**

**M.S.** **Applied Data Science (3.85 / 4.00) *Dec 2019***

***School of Information Studies (iSchool), Syracuse University, Syracuse, New York***

Coursework: Big Data Analysis, Scripting for Data Analysis, Data Analysis and Decision Making, Data Analytics, Database Management, Business Analytics, Marketing Analytics, Data Science, Data Visualization, Data Warehousing.

**B. Tech.** **Computer Science and Engineering *May 2017***

***Walchand College of Engineering Sangli, Maharashtra, India***

Course Work: Machine Learning, Business Intelligence, Data Mining, Advanced Data Structure.

**PROFESSIONAL EXPERIENCE:**

**Data Science Intern: Gartner Inc. Arlington, Virginia *May 2019 - Aug 2019***

* Developed a python package for text processing to facilitate text cleaning, TFIDF and Collocation analysis, Visualization, LDA, Labeled LDA and Clustering.
* Researched and implemented multiple research papers by leveraging IEEE articles and numerous websites with related papers to determine optimized techniques for each feature implemented in the package.
* Manipulated Random Forest technique with personal functions to traverse the trees for finding feature importance of 24 labels.
* Used CountVectorizer and Gensim Word2Vec embeddings with Random Forest to predict 96 labels of 500K rows of text data.

**Data Analyst: iConsult Collaborative Syracuse, New York *Mar 2019 - Present***

* Identified Key Performance Indicators (KPIs) for NEXIS and formulated strategies to increase website traffic by 20% and tracked the increase in online presence of users using Google Analytics and validated the recommendations.
* Analyzed the customer data for NEXIS using R and python and created dashboards in Tableau, and Power BI.
* Leveraged Jira to collaborate and concur on project ideas, live tracking of projects and internal document management.

**ACADEMIC PROJECTS:**

**Data Scientist: Fare prediction of NYC Taxies  *Aug 2018 - Dec 2018***

* Extracted the NYC Yellow cab data from Kaggle and analyzed it to determine the most profitable duration and fair for a ride.
* Used PySpark and XGBoost to model the data and applied PCA to find the best variables to stop overfitting of data.
* Created a model comprising of Random Forest, Gradient Boosting and RMSE to learn the error in prediction.
* Applied XGBoost to obtain an accuracy of 67.7% for ride duration and 63% for fair.

**Neural Programmer: Program Induction using Neural Program Interpreter *Jan 2019 - Apr 2019***

* Built a Neural Program Interpreter using program induction to automate priority sorting, arithmetic question answering and card matching together with other machine learning tasks.
* Used a trained LSTM as a Neural Program Interpreter to learn new programs storing the already learned programs.
* Employed a novel architecture, which consisted of three different components, for the LSTM to learn new programs.

**Data Scientist: Recommendations for a Hotel Group *Jan 2018 - Apr 2018***

* Performed data cleaning, descriptive statistics, transformation, analysis and visualization on over 10.5 million data points using R.
* Provided strategies to the hotel group based on the actionable insights generated by focusing on the Net Promoter Score (NPS).
* Applied Linear Regression, Apriori, SVM and Association Rules to find and support rules of the recommendations.

**Scripting Data Analyst: Twitter Sentiment Analysis *Jan 2018 - Apr 2018***

* Analyzed tweets of soccer clubs to determine which club had the most loyal fan following by acquired 10000 semi-structured tweets from twitter accounts of 4 renowned soccer teams.
* Employed PyMongo to collect the tweets and place them in a MongoDB repository while utilizing Python for analysis.
* Derived a calculated score using likes, retweets, fan following and tweets per match day to determine Fan Loyalty.

**TECHNICAL SKILLS:**

**Languages & Database:** C/C++, Java, Python, R, MySQL, Access, MongoDB, CouchDB, Oracle, MS SQL, SAP.

**Tools & Software:** Jupyter Notebook, Minitab, Excel, Tableau, RShiny, Docker, Hadoop, Weka, Jira, Putty, Winscp, AWS, SAS.

**Machine Learning:** Regression, SVM, PCA, Random Forest, Decision Tree, Gradient Boosting, XGBoost.

**Deep Learning:** tensorflow, keras, Convolutional/ Recurrent/ Wide residual/ Echo state/ Hopfield Neural Networks.

**Packages:** pandas, numpy, matplotlib, scipy, scikit-learn, NLTK, genism, spacy, pyspark, seaborn, ggplot2, dplyr.

**LEADERSHIP:**

* Leader: charity cricket match in Sept. 2018 & Jan. 2019 for Nanhi Kali Syracuse, which helps girl child in education.
* Captain: district’s Cricket team in U-14 and U-16 to reach the district finals twice in 4 years.
* Captain: Emmanuel English School’s cricket team to win 2 interschool championship.