Jay Kachhadia

jaykachhadia.me | Data Engineering Intern @ ViacomCBS | Writer - Towards Data Science | Tableau Desktop Specialist®

+1 332 201 7120
kachhadiajay@gmail.com
linkedin.com/in/jkachhadia

Education

Syracuse University, School of Information Studies — May 2020
MS Applied Data Science | GPA: 3.77

S.V. National Institute of Technology (NIT), India — May 2018
Bachelor of Technology in Computer Engineering | Distinction Holder

Experience

CBS Interactive, a ViacomCBS Company — Jun 2019 - Aug 2019
Data Engineering Intern - Data & Insights Group

- Built an Anomaly Detection pipeline which ingests KPIs of 46 multi-platform CBS properties daily and deployed it on GCP
- Engineered features and a time-series forecasting model outdoing previous system(Adobe Sensei)'s accuracy by 18.26%
- Developed an approach to represent 30 million user devices and it's content viewing patterns on CBS All Access on graphs
- Designed an unsupervised clustering algorithm for node embeddings to detect subscription abuse (password sharing)

Center for Computational and Data Science, Syracuse University — Aug 2018 - Present
*Research Assistant under **Prof. Jennifer Stromer-Galley***

- Performed text mining to discover patterns in 18 million public comments of presidential elections using machine learning
- Developed NLP algorithms for social media data classification of 1100 candidates in US elections of 2018 and 2020
- Performed space and time efficient transformations with migration into MongoDB for more than 80 million data points
- Building streaming data pipelines in Apache Spark and Airflow for ingesting Twitter, Facebook and Instagram data
- Implemented on-the-go preprocessing with online machine learning models in streaming data pipelines [website]

Indian Institute for Human Settlements — May 2017 - Jul 2017
Data Science Intern - Geospatial and Urban Informatics Lab

- Automated Land cover classification process of Landsat 8 satellite Imagery into Land Use Land Cover Maps using python
- Increased classification accuracy to 92.44% using machine learning compared to traditional method's 75-80% accuracy
- Packed the entire automation process into a PyQt application with a feedback system for the classification algorithm
- Used by the United Nations Development Program for planning 7 sustainable cities in India [research paper]

Projects

C-REX, A Graph-based Recommendation Engine | Python | Neo4j | Machine Learning | *HACKDAY 2019 Winner at CBS*

- Built a graph-based hyper-personalized recommendation engine connecting 3rd Party shows with CBS shows
- Developed a multi-dimensional recommendation space for shows and user personas using Neural Networks(Node2Vec)

Big Data, Big Money | Python | Spark | Keras | Data Science

- Performed Time Series Analysis and modeling on 7 million records of 6000 stocks of US markets using Spark
- Developed a LSTM time-series model to predict investment returns for making calculated decisions involving minimal risk

Indian Premier League Game Predictions | Python | Dash | Data Science

- Performed feature engineering to generate relevant attributed from 179 thousand ball-by-ball data points of 756 games
- Developed a Dash application powered by machine learning models to predict scores and win probabilities [webapp link]

*All the public projects are available on https://github.com/jkachhadia

Publications & Presentations

[1] Gupta S., **Kachhadia J.**, Bolden S., Korsunska A., and Stromer-Galley J., "**Understanding Political Discourse on Social Media using BERT**", 6th International Conference on Computational Social Science, **MIT**, Cambridge, MA, July 2020

[2] Gupta S., **Kachhadia J.**, Bolden S., Korsunska A., and Stromer-Galley J., "**PoliBERT: Classifying political social media messages with BERT**", 2020 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation, Washington, D.C., October 2020

Community Engagements

Writer | Towards Data Science | Medium Publication | https://medium.com/@jkachhadia — Jun 2017 - Present

- Write about my journey in Data Science to contribute to the community of 392K with 21,000+ readers of my own blogs

Technology & Software Skills

Programming Languages	Python | Scala | Cypher | SQL | CloudSQL | HTML | CSS | Javascript | R | git
Frameworks	Flask | Apache Spark | Apache Airflow | PyQt | Material Design | Chart.js | PyTorch | Keras
Libraries	Scikit Learn | MLLib | NLTK | Node2Vec | Word2Vec | OpenCV | Pandas | Numpy | Matplotlib
Softwares	Tableau | Excel | Photoshop | Edge Animate | After Effects | Premiere
Databases	Neo4j | MongoDB | MySQL | Firebase | RethinkDB | BigQuery