Shashank Kapadia
Data Science Manager

shmkapadia@gmail.com ✉

+1 (617) 794 3430 📱

Boston MA 02135 📍

github.com/kapadias 🔵

WORK EXPERIENCE

Manager, Data Science
Monster Worldwide

05/2019 – Present   *Boston, MA*

- Manage a cross-functional team of 4 data scientists and 2 engineers to lead the development of personalization services reshaping core recommender system driving >150MM searches annually
- Implement BERT based NLP pipeline for Named Entity Recognition and Relation Extraction using >1MM parsed resume data to remodel internal knowledge graph powering 20+ product features
- Spearhead the development of a multitude of statistical models ranging from CTR prediction, LTV forecasting, market/individual level response modeling, and targeting optimization influencing >$50MM in ads revenue
- Define the trajectory and the growth of the data science team, including forming and mentoring a team and streamlining Machine Learning Operations (MLOps) processes on cloud reducing time to delivery by >30%

Senior Data Scientist
Digitas Inc.

10/2016 – 05/2019   *Boston, MA*

- Headed the 3 data scientists' team, working alongside engineering to develop Customer Data Platform, enriching internal data, and optimizing $200MM+ marketing budget through user-behavior modeling, clustering, and performance analytics
- Delivered an XGBoost, and Random Forest based classification model for scoring >40MM customers propensity to purchase within a product category resulting in more compelling user engagement and improving previous results by 12.5%
- Orchestrated the power of data, machine learning, and story-telling to develop an interactive dashboard capturing US mobile phone market trends by analyzing >300MM daily transactions and transforming into 40+ features

Model Validation Engineer II
Air Worldwide

02/2015 – 10/2016   *Boston, MA*

- Developed DBSCAN based spatial clustering model to provide better insights and resolution into insurance claims exceeding loss estimation accuracy by >30%
- Automated data exploration and feature-engineering processes across disparate data sources by wrangling >20MM data points using PySpark and crontab reducing turnaround reporting time by >50%

SKILLS

Programming & Databases
Python, R, SQL, MySQL, Hive, PostgreSQL, Hadoop, Google BigQuery

Packages & Frameworks
Pandas, Numpy, Scikit-Learn, TensorFlow, PyTorch, Keras, NLTK, Gensim, Spacy, PySpark

Infrastructure and Tools
Google Cloud Platform (GCP), Amazon Web Services (AWS), Apache AirFlow, JIRA, Git, Tableau, Google Data Studio

PUBLICATIONS

Journal Article
Determining the optimal collection period for returned products in a stochastic environment ⬈

Author(s)
Nizar Zaarour; Emanuel Melachrinoudis; Shashank Kapadia; Hokey Min

May 10, 2019

ARTICLES

Recommendation System in Python: LightFM ⬈
A guide to building a recommender system in Python using LightFM

Evaluate Topic Models: Latent Dirichlet Allocation ⬈
A guide to building optimal and interpretable topic models in Python

Language Models: N-Gram ⬈
An introduction and step into statistical language modeling

AWARDS AND HONORS

Winner, Data and Analytics Challenge (06/2018) ⬈
Publicis Groupe
- The cross-agencies data and analytics competition for developing Word2Vec based NLP models on 10K+ physician-patient conversations, and visualizing actionable insights to enhance patients experience

Master Valedictorian (05/2015)
Northeastern University
- The Top Graduate Award recognizes Master of Science students who, along with graduating at the top of their class, have performed with persistence and consistency throughout their degree program

EDUCATION

MS, Operations Research
Northeastern University

09/2012 – 12/2014   *Boston, MA*

BS, Electronics & Communication
University of Pune

08/2008 – 05/2012   *Pune, India*