

Manoj Kumar

☎ (201-899-0725) | ✉ mks542@nyu.edu | 📱 MechCoder

Education

New York University

M.S. IN DATA SCIENCE - 3.84/4.0

New York City

Sept. 2015 - PRESENT

Birla Institute of Technology and Science

B.E. IN MECHANICAL ENGINEERING - 8.62/10.0

Goa, India

Aug 2011 - May 2015

Open-source development

Scikit-optimize

CO-CREATOR

June 2016 - Present

- Co-authored an easy-to use sequential model-based optimization toolkit supporting Gaussian Process and Random Forest based search

Scikit-learn

CORE DEVELOPER

2013 - Present

- Co-mentor – Google Summer of Code 2016, Cython-fused types support
- Student – Google Summer of Code 2014, Improved speed and memory of linear models

Apache Spark

GOOGLE SUMMER OF CODE 2015, STUDENT

Remote

May 2015 - Aug 2015

- Improved PySpark's machine learning capabilities by writing wrappers around existing algorithms in Scala
- Wrote the runWithValidation utility in GradientBoostedTrees to support early-stopping
- Added support for sparse input and ability to load and save GaussianMixtureModels

SymPy

GOOGLE SUMMER OF CODE 2013, STUDENT

Remote

May 2013 - Aug 2013

- Improved symbolic differential equation solving support
- Wrote the abstractions for solving partial differential equations

Experience

IBM Spark Technology Center

SOFTWARE ENGINEERING INTERN

San Francisco

June 2016 - Sept 2016

- Improved support for distributed benchmarking and speed of MLlib's Random Forest implementation upto 10%
- Wrote a Python-API for Apache SystemML's MLContext using Py4j

New York University

JUNIOR DATA SCIENTIST, SCIKIT-LEARN

New York City

Oct 2015 - May 2016

- Assisted in the 0.17 release by bug-fixing and code-reviews
- Implemented HuberRegressor, a linear model that is robust to outliers
- Implemented SelectFromModel, a meta-transformer to search for prediction thresholds

Telecom ParisTech

RESEARCH ENGINEER, SCIKIT-LEARN

Paris

July 2014 - Dec 2014

- Implemented Birch, an online clustering algorithm as an alternative to KMeans
- Added support for Multinomial Logistic Regression by polishing the existing implementation

Projects

EmineT: A variational LSTM to generate rap

INFERENCE AND REPRESENTATION

Oct 2016 - Dec 2016

- Used a LSTM to model the distribution of amplitudes of the raw waveform at a given time-step given the previous time-steps
- In addition, used a variational auto-encoder to model the latent variables at every time-step of the LSTM
- Time-aligned the audio with lyrics, to provide phonemes as input in combination with the raw waveforms

Reading Comprehension

NATURAL LANGUAGE UNDERSTANDING WITH DEEP LEARNING

Oct 2016 - Dec 2016

- Implemented the Stanford attention-based Reader to do question answering on the Who-did-What dataset in tensorflow
- Implemented variants of the reader to incorporate information based on choices

Taxi-News

BIG DATA

Feb 2016 - May 2016

- Analyzed the effect of taxi-related news events such as Uber-ban on yellow cab ridership
- Wrote an efficient map-reduce implementation using the rtree module to find the number of taxis in a given NYC neighbourhood for a given date

HitOrFlop

BUSINESS ANALYTICS

Oct 2015 - Dec 2015

- Predict if ratio between opening weekend revenue and the budget is greater than a user-defined threshold using pre-released data mined from iMDB
- Determined the features most predictive of a movie's success

Technical Strengths

PROGRAMMING

- Python, Scala, C, Octave