publications (selected)


Software

**pomegranate** (2,299 stars, 425 forks, used by 215 repos, 565,502 downloads (PePY) as of 5/5/20)
pomegranate is a Python package for probabilistic modeling. It extends scikit-learn by offering a more flexible API for building and training complex probabilistic models, such as Bayesian networks, hidden Markov models, and mixture models. Users can build models with the many pre-defined distributions or easily implement their own custom ones. [https://github.com/jmschrei/pomegranate](https://github.com/jmschrei/pomegranate)

**apricot** (264 stars, 27 forks, 19,505 downloads (PePY) as of 5/5/20)
apricot is a Python package that implements submodular optimization for the purpose of summarizing massive data sets into non-redundant subsets that still represent the space of the full data. The package follow the format of scikit-learn so that selection can be done easily and without background knowledge and dropped into existing pipelines. [https://github.com/jmschrei/apricot](https://github.com/jmschrei/apricot)

**Avocado** (68 stars, 11 forks, 20,210 downloads (PePY) as of 5/5/20)
Avocado is a Python package that implements deep tensor factorization for the purpose of modeling large, but incomplete, compendia of epigenomic data. The model both learns a low-dimensional representation that is broadly useful and can be used to impute the missing values in the tensor. [https://github.com/jmschrei/avocado](https://github.com/jmschrei/avocado)

**scikit-learn** (>40.5k stars, >19.6k forks, used by >105k repos, >198M downloads as of 5/5/20)
sckit-learn is a Python package that implements classic supervised and unsupervised machine learning algorithms as well as many components of the machine learning ecosystem, such as model evaluation, hyperparameter selection, and data preprocessing steps. I contributed for several years and was a core contributor for around a year, focusing on the tree-based methods (specifically gradient boosting) and probabilistic models. I am now an emeritus core developer because I do not regularly contribute right now. [https://github.com/scikit-learn/scikit-learn](https://github.com/scikit-learn/scikit-learn)

Talks

**Avocado**  Learns a Latent Representation of the Human Epigenome
**Rambutan** Predicts Three Dimensional Structure of the Human Genome
**pomegranate**: probabilistic modeling in python
**apricot**: submodular optimization for machine learning

scipy (2019), Moore-Sloan Data Science Summit (2019)