

# Computational Linguistics

---

## Rules and Statistics

Kenneth Steimel



# What is scikit-learn?

- High level machine learning api
  - Gathers together a bunch of machine learning algorithms, feature selection algorithms, preprocessing procedures
  - Provides an easy way to extend the prebuilt models with your own
  - Construction of whole Machine Learning pipelines.
- Most popular machine learning toolkit in python
- Strictly non-sequence models
- Built on the numpy and scipy packages
- Documentation available at <https://scikit-learn.org>

# How do you use scikit-learn?

- Within python or ipython
- Instalation
  - Windows `pip install scikit-learn` or via Anaconda
  - MacOS and linux should probably search for sklearn in your package manager

## API design

- Components are initialized before learning is done.
  - `model = LinearSVC()`
- Scikit-learn components do three main things
  - Fit
  - Predict
  - Transform
- Scikit-learn uses duck typing for this api. No need to inherit from a class.

## Hands on demo

- Let's see some examples so this API design makes a bit more sense.