Statistical Forecasting Internship
Vaisala Inc.

Anna Sinilo
September 12, 2016
Agenda

- How to find an internship: tips
- How to pass an interview: tips
- About Vaisala and team I worked for
- My project
- Statistical techniques I used
- Tools and technologies I used
- My coursework
How to find an internship: tips

- What worked for me:
  - Different [data science] meet-ups (broadens mind)
  - Career Center at SJSU (resume review, spartan jobs, on-campus interviews, talks)
  - Go directly to corporate websites
  - Tailor your resume to every application (quality matters a lot!)
  - Have a spreadsheet with application info to control your search

- What didn’t work:
  - LinkedIn and other job boards (too competitive)
  - Staffing agencies
How to pass an interview: tips

What worked for me:
- Cheat sheet with common questions and my own answers
- Know general info (about company, interviewer)
- Be confident and “keep talking”
- Show motivation and how you fit to this role

What didn’t work:
- Asking to repeat/or explain the question
- Giving long answers
About Vaisala Inc.

- Headquarters in Finland, Helsinki
- Develops, manufactures and markets products and services for environmental and industrial measurement (sensors, radiosondes, radars, etc.)
- Operates in 140 countries
- Celebrates its 80th year of operation
- Over 1500 employees
- Giant Leap Internship Program (~25 interns every summer)
- Requirements for my project: student in statistics, knowledge of forecasting models, python/matlab
My office and team

- Located in Seattle, WA
- Office includes forecasting and assessment teams
- I worked in Forecasting:
  - People in my team:
    - 5 people (not including me)
    - 2 with background related to Meteorology, 2 – Atmospheric Science, 1 – Statistics
- Produce forecasts and other consulting services to clients (owners of wind/solar sites, airports, etc.)
My project:

- seasonal forecasting of wind and solar energy generation on multi-month ahead horizons
Statistical techniques I used

- Exploratory data analysis:
  - Basic statistical measures: mean, std/var, correlations, etc.
  - Plots

- Feature engineering

- Dimensionality reduction (PCA)

- Missing values and outlier analysis
Plots

Subset of predictors (overall nearly 300 predictors)

4 types of response
**Statistical models I used**

- **Regression**
- **Model evaluation:**
  - MAE, MSE, RMSE, L-norms

- **Classification (2-class)**
- **Model evaluation:**
  - accuracy, precision, recall, area under the ROC curve

- **Time series**
- **Model evaluation:**
  - MAE, MSE, RMSE, L-norms
Tools and technologies I used

- Python (Ipython/Jupyter Notebook, PyCharm IDE)
  - Packages: numpy, scipy, scikit-learn, pandas, statsmodels, seaborn, etc.

- R (RStudio IDE)
  - Packages: quantmod, ggplot2, etc.
My coursework

- Classes I recommend:
  - Math163 Probability Theory
  - Math261a Regression Analysis
  - Math265 Time Series Analysis
  - Any machine learning/data mining class (possibly online)
  - Kaggle competitions

- Finding:
  - Learning a particular technology – manageable
  - Learning how to interpret output, implement new concepts – much harder
Thank you for your attention!