COMP 541 DATA MINING | PROJECT 6

TRAVNUR

Improving Healthcare Through Data Mining

Verab Chitchyan
Gabriella de Asis
Emmanoel Dermkrdichyan
Carson Logston
Anh (Steven) Nguyen

MODELING TECHNIQUE

- Random Forest
  - classification of staffing shortage vs. anticipated shortage
  - initial 33% accuracy
- Linear Regression
  - increase in confirmed COVID patients result in higher staff shortages
  - initial 52%+ accuracy

- sklearn
  - linear_model
  - RandomForestClassifier

TESTING

- Training set
  - 80% of the dataset, 11,392 rows
- Test set
  - 20% of the dataset, 2,849 rows

- sklearn.model_selection
  - test_train_split
  - model.fit(X_train, Y_train)
- Y_pred = model.predict(X_test)
  - r2_score(Y_test, Y_pred)

PARAMETERS

Target value: Critical staffing shortage today

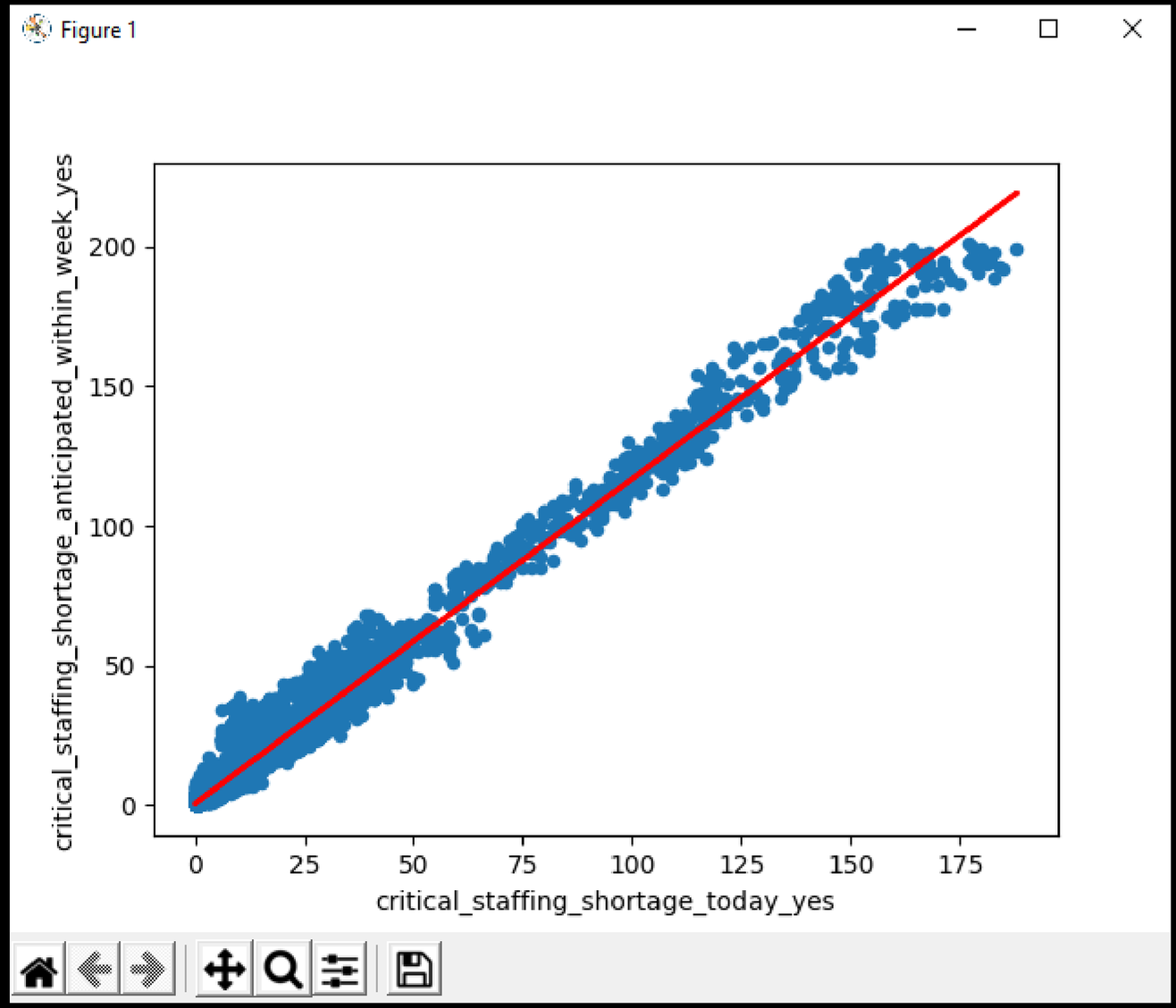
3 sets of parameters, 3 linear regression models

- Model No. 1
  - staff shortage, confirmed adult cases, and confirmed pediatric cases
- Model No. 2
  - staff shortage, confirmed adult cases, and confirmed pediatric cases
  - anticipated shortage
- Model No. 3
  - staff shortage, confirmed adult cases, and confirmed pediatric cases
  - anticipated shortage, confirmed and suspected adult cases
  - confirmed and suspected pediatric cases

RESULTS

- Model No. 1
  - 46.7% accuracy
- Model No. 2
  - 83.7% accuracy
- Model No. 3
  - 83.1% accuracy

Red line:
linear regression
prediction model



MODEL RANKING

- Model No. 2 is the most accurate
  - only confirmed cases
  - hospital anticipated shortage
- Model No. 3 is a close second
  - 0.6% difference is insignificant
  - not all suspected COVID cases are actual COVID cases
- Model No. 1 is not accurate
  - does not factor in other attributes such as
    - other illnesses, funding, etc.
  - each hospital's anticipated shortages increases prediction accuracy by roughly 37%

ADDITIONAL DATASET

- CA Nurse Shortage dataset
  - the full dataset became inaccessible
  - only one day was provided
  - shortage severity classified as low, medium, and high
- Random forest classification model
  - 83% accuracy
  - only for that one day
  - this machine learning algorithm is best suited towards classification, and the prediction model turned out far better for the CA nurse shortages dataset than our other one.

GOAL SATISFACTION

- Business Objectives
  - provide travel nurses to hospitals with staffing shortages
- Success Criteria
  - predict staffing shortages within 1 week in advance with 70% accuracy or higher
- Our prediction model
  - is able to predict staffing shortages 1 week in advance with an estimated 83% accuracy