Predicting Women’s Freedom

Worldwide Safety and Security of Women

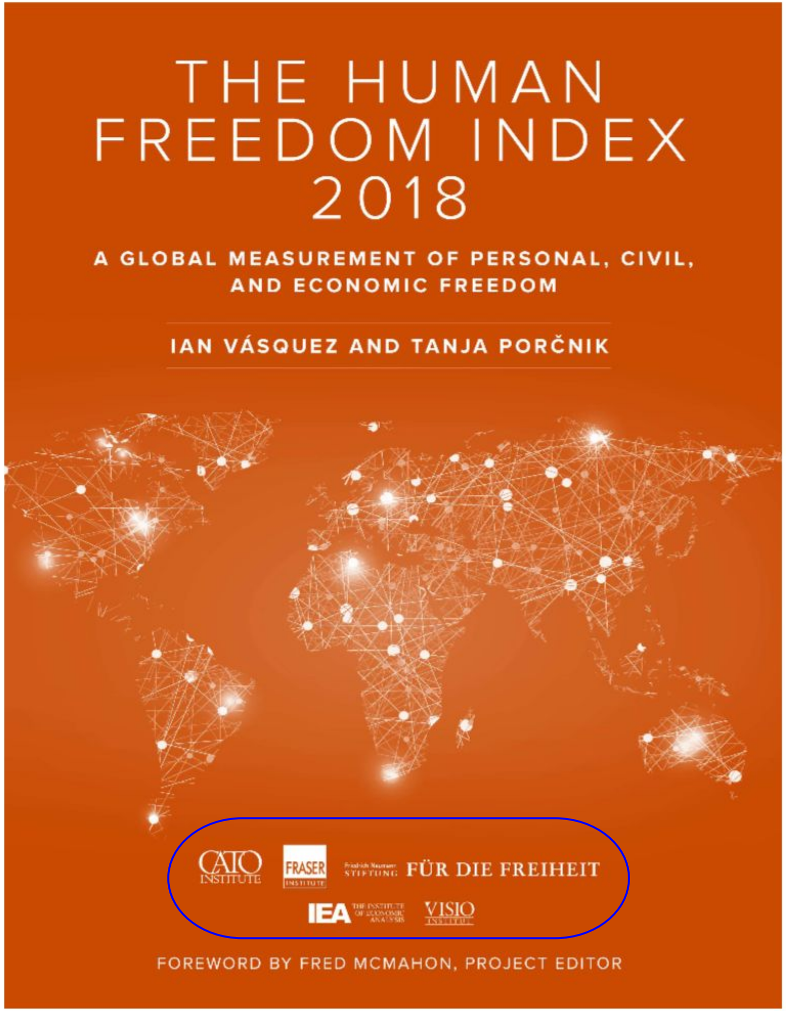
The Data

Personal, civil, & economic freedoms

Cato Institute, the Fraser Institute, and the Liberales Institut

162 countries

2008 - 2016 and 4 publications



The Data

The Human Freedom Index presents a broad measure of human freedom, understood as the absence of coercive constraint. It uses 79 distinct indicators of personal and economic freedom.

Scale: 0 to 10

Missingness:

50%

1458 => 1127

Table 1. Structure of the Human Freedom Index

PERSONAL FREEDOM

1. LEGAL PROTECTION AND SECURITY

- A. Rule of Law
  - i. Procedural Justice
  - ii. Civil Justice
  - iii. Criminal Justice
- B. Security and Safety
  - i. Homicide
  - ii. Disappearances, Conflicts, and Terrorism
    - a. Disappearances
    - b. Violent Conflicts
    - c. Organized Conflicts
    - d. Terrorism Fatalities
    - e. Terrorism Injuries
  - iii. Women's Security and Safety
    - a. Female Genital Mutilation
    - b. Missing Women
    - c. Inheritance Rights
      - Widows
      - Daughters

2. SPECIFIC PERSONAL FREEDOMS

- A. Movement
  - i. Domestic Movement
  - ii. Foreign Movement
  - iii. Women's Movement
- B. Religion
  - i. Establishing and Operating Religious Organizations
  - ii. Harassment and Physical Hostilities
  - iii. Legal and Regulatory Restrictions
- C. Association, Assembly, and Civil Society
  - i. Association
  - ii. Assembly
  - iii. Establishing and Operating Political Parties
  - iv. Establishing and Operating Professional Organizations
  - v. Establishing and Operating Educational, Sporting, and Cultural Organizations
- D. Expression and Information
  - i. Press Killed
  - ii. Press Jailed
  - iii. Laws and Regulations That Influence Media Content
  - iv. Political Pressures and Controls on Media Content
  - v. Access to Cable/Satellite
  - vi. Access to Foreign Newspapers
  - vii. State Control over Internet Access
- E. Identity and Relationships
  - i. Legal Gender
  - ii. Parental Rights
    - a. In Marriage
    - b. After Divorce
  - iii. Same-Sex Relationships
    - a. Male-to-Male Relationships
    - b. Female-to-Female Relationships
  - iv. Divorce

ECONOMIC FREEDOM

- A. Size of Government
  - i. Government Consumption
  - ii. Transfers and Subsidies
  - iii. Government Enterprises and Investments
  - iv. Top Marginal Tax Rate
    - a. Top Marginal Income Tax Rate
    - b. Top Marginal Income and Payroll Tax Rate
- B. Legal System and Property Rights
  - i. Judicial Independence
  - ii. Impartial Courts
  - iii. Protection of Property Rights
  - iv. Military Interference in Rule of Law and Politics
  - v. Integrity of the Legal System
  - vi. Legal Enforcement of Contracts
  - vii. Regulatory Restrictions on the Sale of Real Property
  - viii. Reliability of Police
  - ix. Business Costs of Crime
- C. Sound Money
  - i. Money Growth
  - ii. Standard Deviation of Inflation
  - iii. Inflation: Most Recent Year
  - iv. Freedom to Own Foreign Currency Bank Account
- D. Freedom to Trade Internationally
  - i. Tariffs
    - a. Revenue from Trade Taxes (% of trade sector)
    - b. Mean Tariff Rate
    - c. Standard Deviation of Tariff Rates
  - ii. Regulatory Trade Barriers
    - a. Nontariff Trade Barriers
    - b. Compliance Costs of Importing and Exporting
  - iii. Black-Market Exchange Rates
  - iv. Controls of the Movement of Capital and People
    - a. Foreign Ownership/Investment Restrictions
    - b. Capital Controls
    - c. Freedom of Foreigners to Visit
- E. Regulation
  - i. Credit Market Regulations
    - a. Ownership of Banks
    - b. Private-Sector Credit
    - c. Interest Rate Controls/Negative Real Interest Rates
  - ii. Labor Market Regulations
    - a. Hiring Regulations and Minimum Wage
    - b. Hiring and Firing Regulations
    - c. Centralized Collective Bargaining
    - d. Hours Regulations
    - e. Mandated Cost of Worker Dismissal
    - f. Conscription
  - iii. Business Regulations
    - a. Administrative Requirements
    - b. Bureaucracy Costs
    - c. Starting a Business
    - d. Extra Payments/Bribes/Favoritism
    - e. Licensing Restrictions
    - f. Cost of Tax Compliance

Question

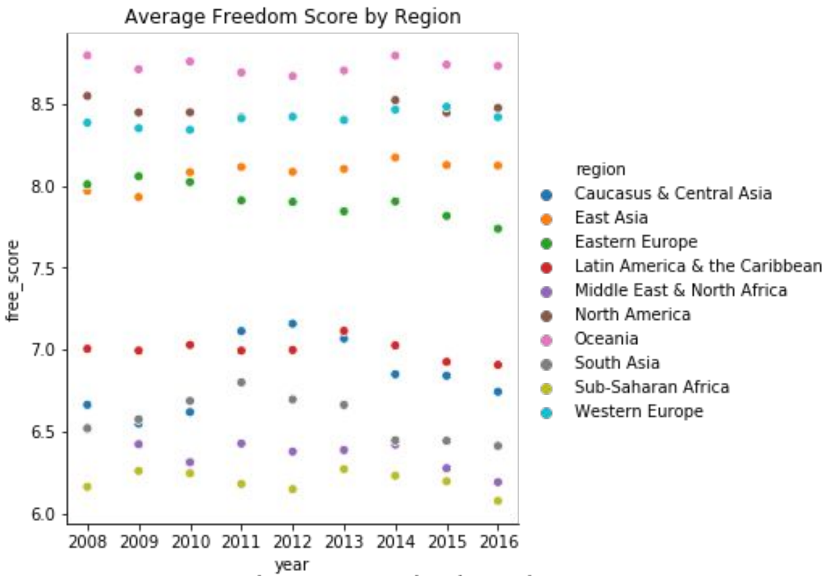
Is it possible to predict if countries will have safer and freer women?

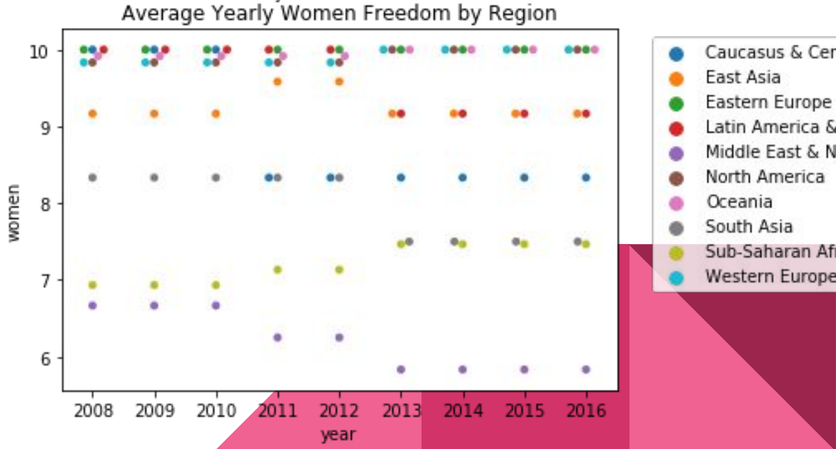
Exploratory Data Analysis

123 => 34

HF Scores by Region

Women Specific





E.D.A.

2008-2016 averages

Free women ≠ Free countries

Overall Freedom

country	
Hong Kong	8.970342
New Zealand	8.857553
Switzerland	8.782947
Canada	8.619081
Australia	8.613201
Finland	8.606345
Denmark	8.589246
Norway	8.527672
United Kingdom	8.513016
Luxembourg	8.494836
Ireland	8.477890
Sweden	8.445823
Austria	8.424141
Netherlands	8.423338
Germany	8.419142
Estonia	8.390369
Malta	8.324871
United States	8.304748
Belgium	8.227984
Czech Rep.	8.209978
Lithuania	8.180404
Japan	8.179346
Portugal	8.165574
Iceland	8.151914
Cyprus	8.148435
Chile	8.144992
Spain	8.141659
Taiwan	8.135682
Mauritius	8.130396
Latvia	8.124512

Name: free_score, dtype: float64

Women's Freedom >>

country	
Netherlands	1.000000
Portugal	1.000000
Korea, South	1.000000
Costa Rica	1.000000
Latvia	1.000000
Colombia	1.000000
Lithuania	1.000000
Luxembourg	1.000000
Malta	1.000000
Canada	1.000000
Mongolia	1.000000
Cambodia	1.000000
Finland	1.000000
New Zealand	1.000000
Bulgaria	1.000000
Norway	1.000000
Panama	1.000000
Kazakhstan	1.000000
Croatia	1.000000
Japan	1.000000
Iceland	1.000000
Estonia	1.000000
Germany	1.000000
El Salvador	1.000000
Guatemala	1.000000
Ecuador	1.000000
Hungary	1.000000
Dominican Rep.	1.000000
Jamaica	1.000000
Denmark	1.000000
Czech Rep.	1.000000
Cyprus	1.000000
Ireland	1.000000
Israel	1.000000
Italy	1.000000
Poland	1.000000
France	1.000000
Spain	1.000000
Belgium	1.000000
Turkey	1.000000
Australia	1.000000
Austria	1.000000
Switzerland	1.000000
United Kingdom	1.000000
Sweden	1.000000
Belarus	1.000000
Slovenia	1.000000
United States	1.000000
Ukraine	1.000000
Slovak Rep.	1.000000
Argentina	1.000000
Romania	1.000000
Russia	1.000000
Venezuela	1.000000
South Africa	0.666667

Name: women, dtype: float64

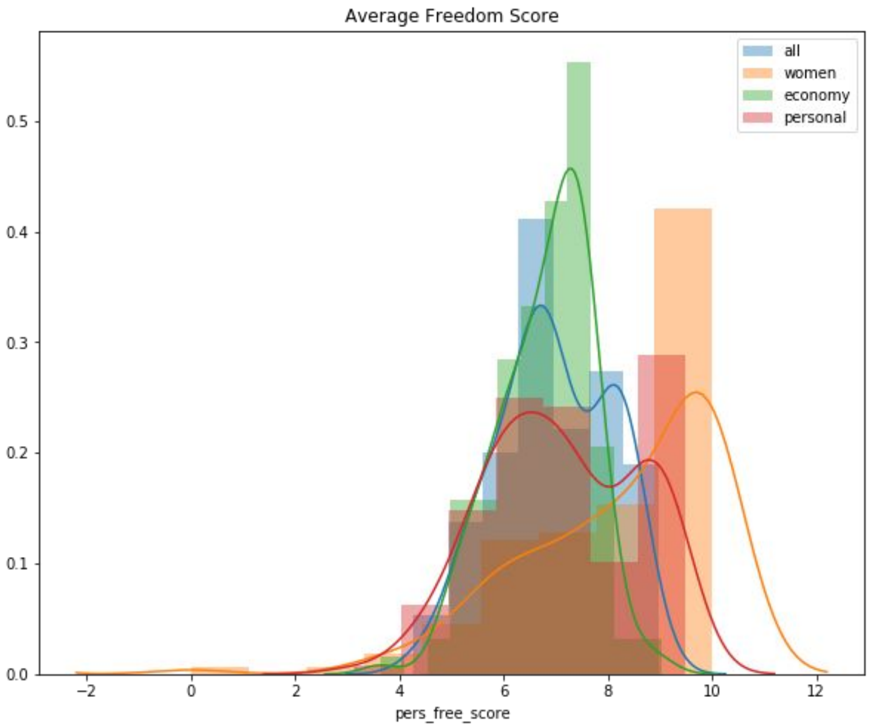
EDA

Women’s Safety and Security

Economic Freedom

Personal Freedom

Overall Freedom

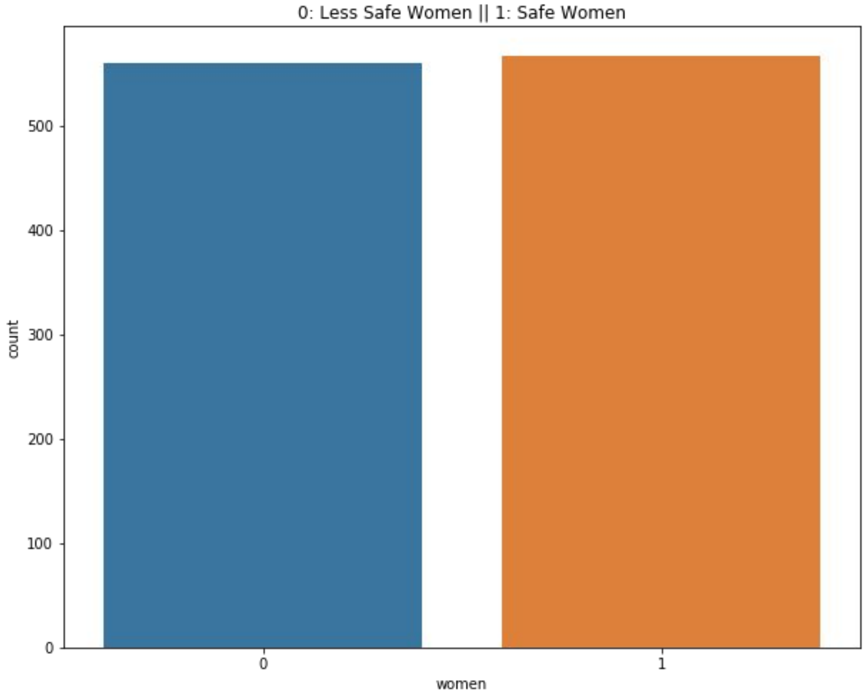


Classification

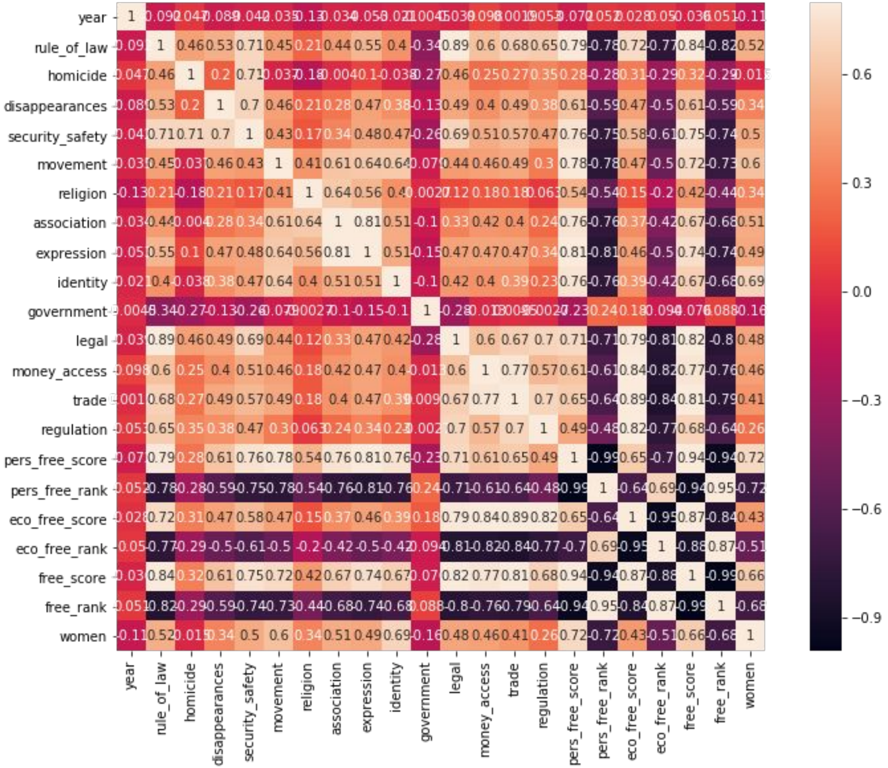
Binary classification

Above / Below Average

Balanced data



Correlation to Target

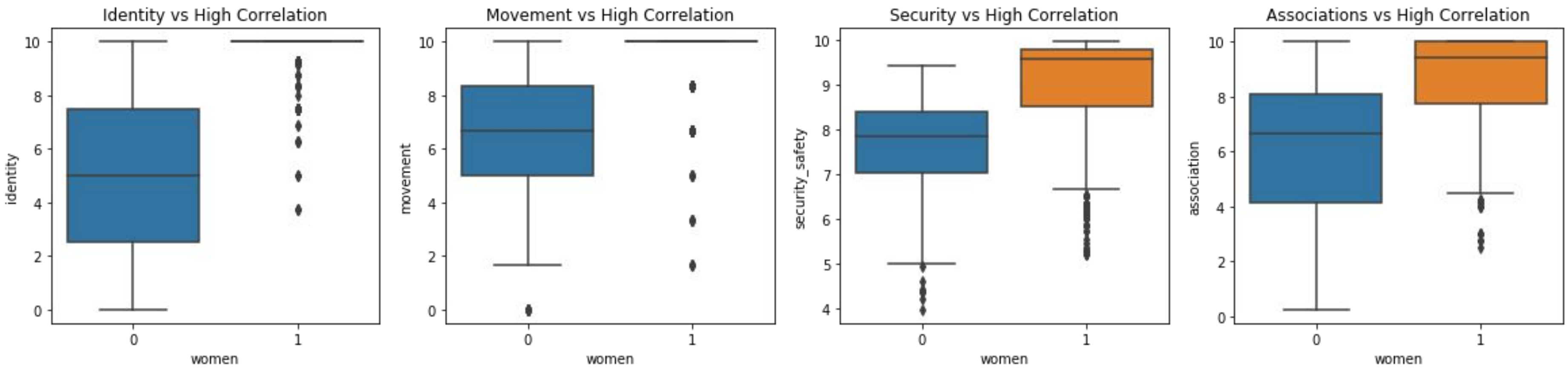


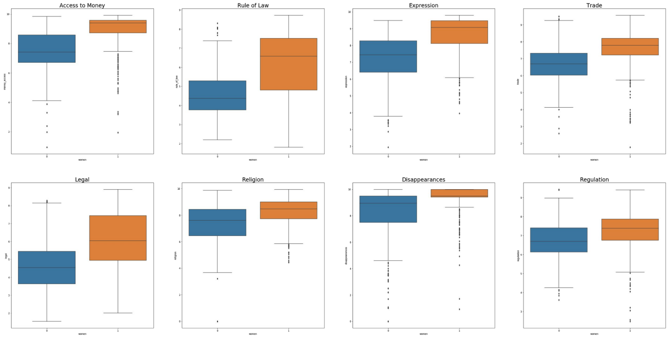
Multicollinearity

```
hfi_df.corr()['women'].abs().sort_values(ascending=False)
```

```
women                1.000000
pers_free_rank       0.715769
pers_free_score      0.715437
identity             0.693493
free_rank            0.680314
free_score           0.659537
movement             0.601298
rule_of_law          0.522270
eco_free_rank        0.510922
association          0.506905
security_safety      0.501125
expression           0.488710
legal                0.480496
money_access         0.460241
eco_free_score       0.429497
trade                0.411200
disappearances       0.343517
religion             0.341972
regulation           0.255533
government           0.158636
year                 0.114688
homicide             0.015391
Name: women, dtype: float64
```

Outliers





Access to money, rule of law, expression, trade, legal, religion, disappearances, & regulation

Classifiers

XGBoost vs.

Logistic Regression

K-Nearest Neighbors

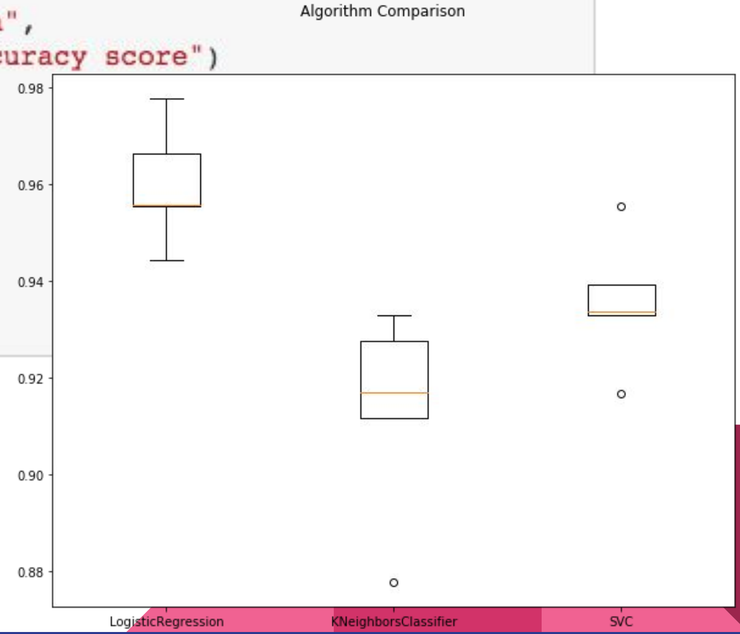
Support Vector Classifier

Training multiple classifiers

```python
results = []
names = []

for key, classifier in classifiers.items():
    classifier.fit(X_train, y_train)
    training_score = cross_val_score(classifier, X_train, y_train, cv=5)
    results.append(training_score)
    names.append(classifier.__class__.__name__)
    print("Classifiers: ", classifier.__class__.__name__, "Has a",
          round(training_score.mean(), 2) * 100, "% training accuracy score")

# boxplot algorithm comparison
fig = plt.figure()
fig.suptitle('Algorithm Comparison')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```



Results

Classifiers: LogisticRegression Has a 96.0 % training accuracy score
Classifiers: KNeighborsClassifier Has a 91.0 % training accuracy score
Classifiers: SVC Has a 94.0 % training accuracy score

GridsearchSV

```
LogisticRegression(C=1000, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='warn',
          n_jobs=None, penalty='l2', random_state=None, solver='warn',
          tol=0.0001, verbose=0, warm_start=False)
LR best Score  0.9944506104328524
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=2, p=2,
           weights='uniform')
KNN best Score  0.9267480577136515
SVC(C=0.5, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma=0.1, kernel='poly',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
 SVC best Score  0.9922308546059934
```

Precision Recall F1

Classification Report: LogR

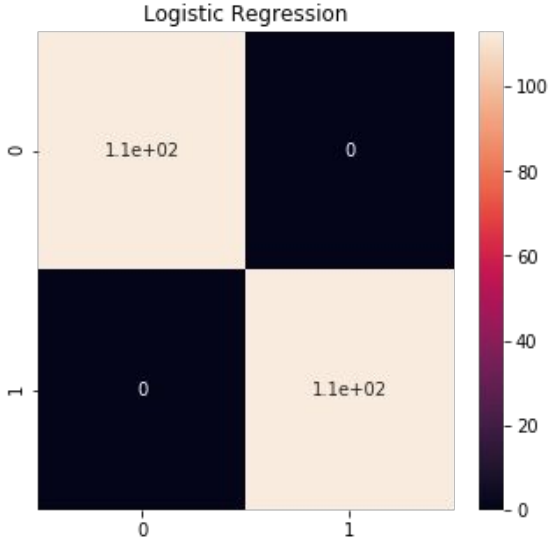
	precision	recall	f1-score	support
0	1.00	1.00	1.00	105
1	1.00	1.00	1.00	121
micro avg	1.00	1.00	1.00	226
macro avg	1.00	1.00	1.00	226
weighted avg	1.00	1.00	1.00	226

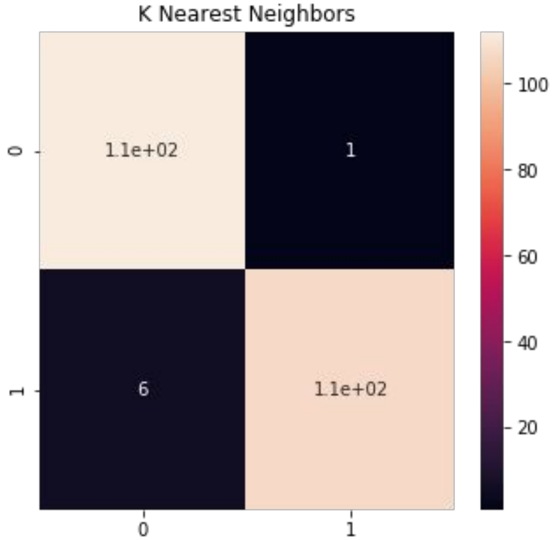
Classification Report: KNN

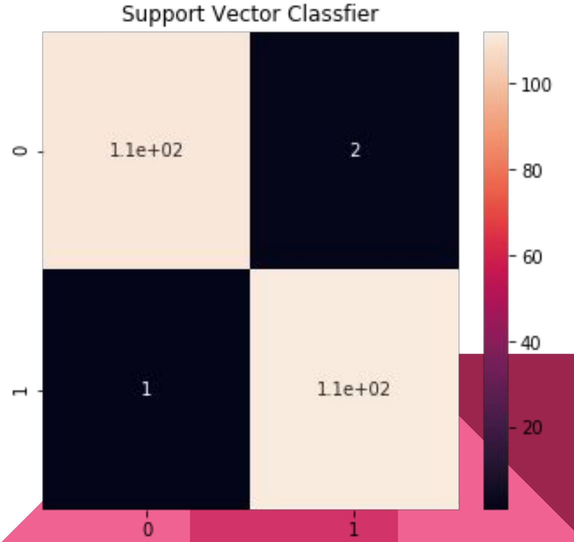
	precision	recall	f1-score	support
0	0.96	0.99	0.98	105
1	0.99	0.97	0.98	121
micro avg	0.98	0.98	0.98	226
macro avg	0.98	0.98	0.98	226
weighted avg	0.98	0.98	0.98	226

Classification Report: SVC

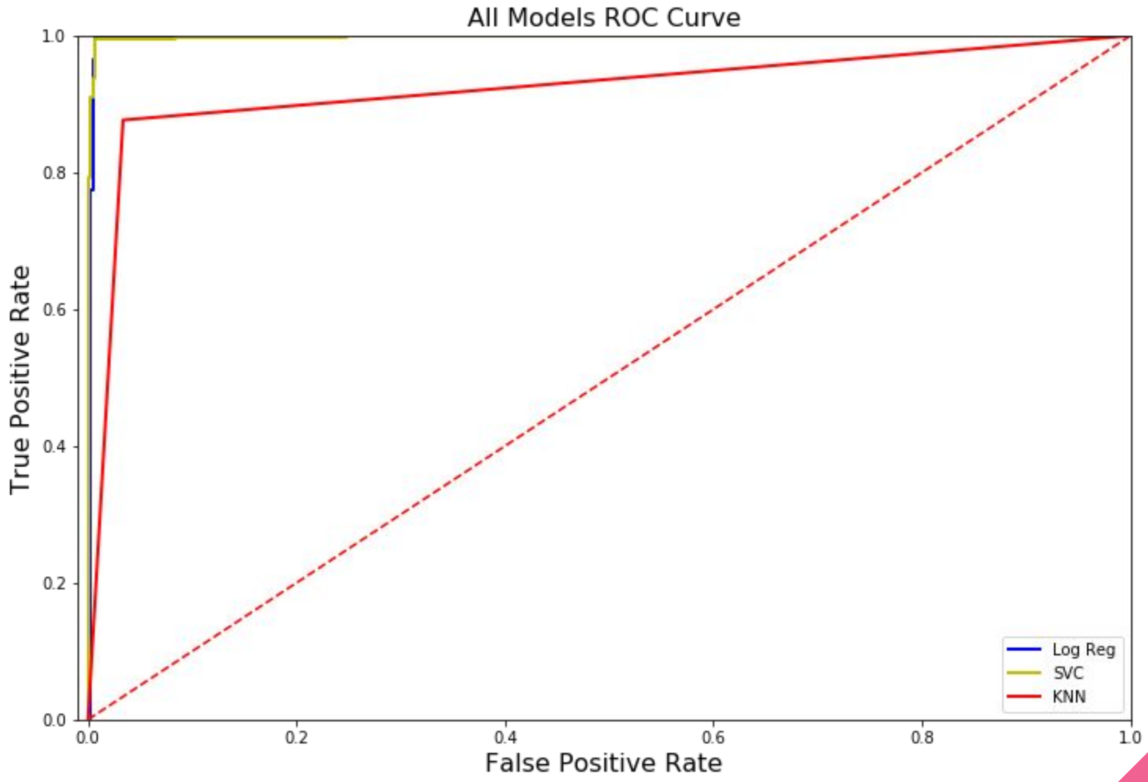
	precision	recall	f1-score	support
0	0.99	1.00	1.00	105
1	1.00	0.99	1.00	121
micro avg	1.00	1.00	1.00	226
macro avg	1.00	1.00	1.00	226
weighted avg	1.00	1.00	1.00	226







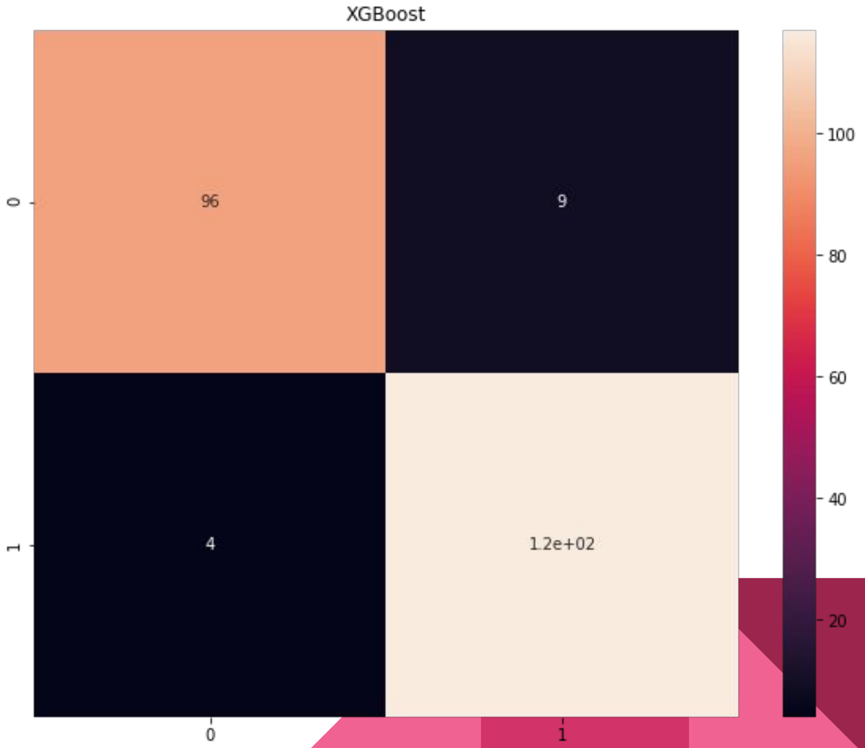
Sensitivity



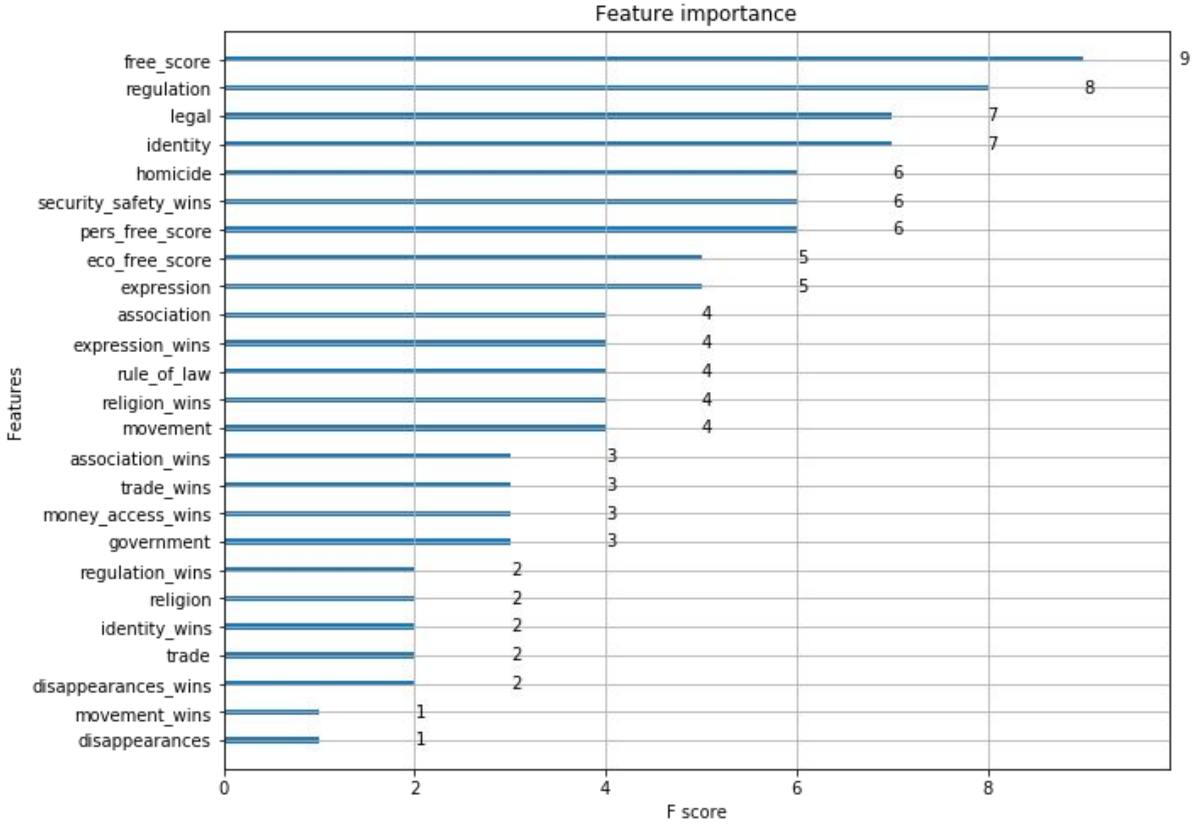
Extreme Gradient Boosting Trees (XGBoost)

```
# alpha = L1 = Lasso, 5 trees deep, 10 trees build
xg_reg = xgb.XGBRegressor(objective ='binary:logistic',
colsample_bytree = 0.3, learning_rate = 0.1,
max_depth = 5, alpha = 10, n_estimators = 10)
```

```
roc_auc_score(y_test, preds)
0.9790633608815428
```



Extreme Gradient Boosting Trees (XGBoost)



roc_auc_score(y_test, preds)

0.9790633608815428

Practical uses

Both logistic regression and XGB worked incredibly well for classification of women's freedom.

Generalized prediction of future data for women

Weak points & shortcomings

World Events

Missing Data

Bias/Overfitting