CS 2750/ISSP 2170
Machine Learning

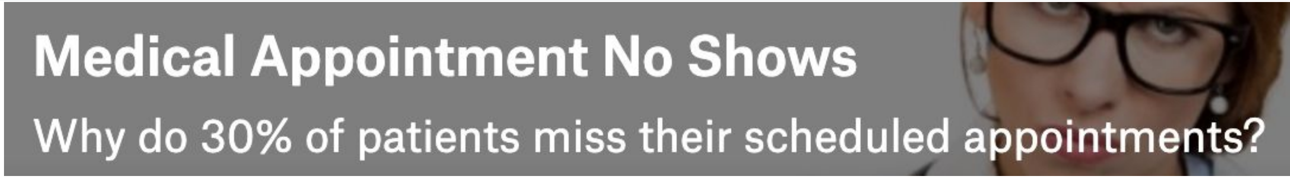
Predicting No-show Medical Appointments

Sanya Bathla Taneja
Intelligent Systems Program
University of Pittsburgh

Outline

- Introduction
- Data
- EDA and Data Processing
- Methods
- Evaluation
- Results
- Conclusion and Future Work

Introduction



What are no-shows?

No-shows are patients who make appointments at a healthcare facility and neither use nor cancel their appointment.

MOTIVATION

- Time
- Money
- Unfilled-slots
- Overbooking

SUGGESTED REASONS FOR NO-SHOW

- Financial concerns
- Long wait times
- Logistic - no transport for appt.
- Forgetfulness

Data: ~21% no-shows

AppointmentID	Unique ID	**Neighborhood**	Unique = 77
PatientID	Unique ID	**Hypertension**	True/False
Gender	Male/Female	**Diabetes**	True/False
Appointment Date	DateTime	**Alcoholism**	True/False
Scheduling Date	Date	**Handicap**	0-4
Age (in years)	-1 to 115	**SMS Received**	Yes/No

110527 instances (patients)

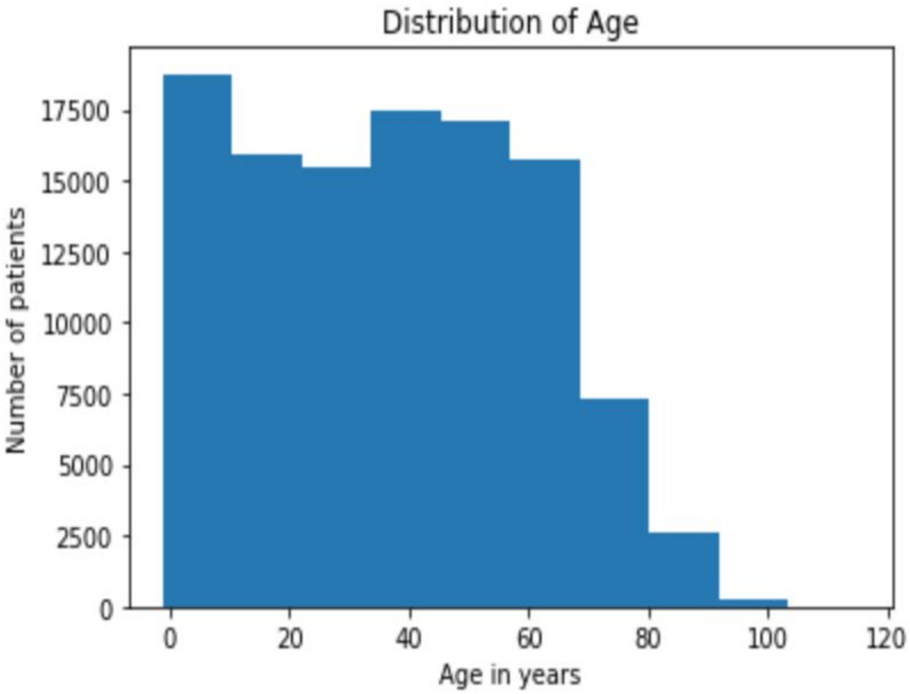
14 variables

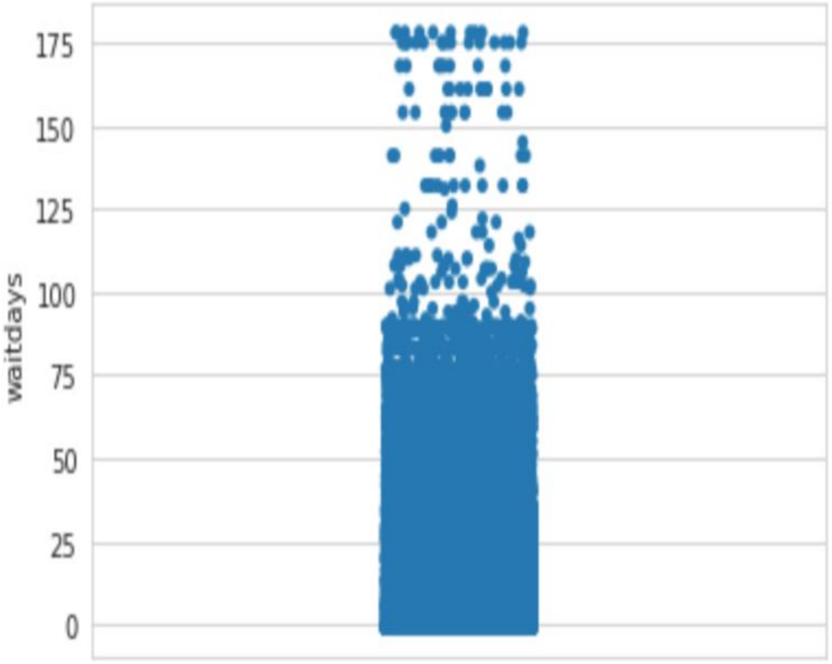
Binary target 'No-show' - Yes/No

Data Processing

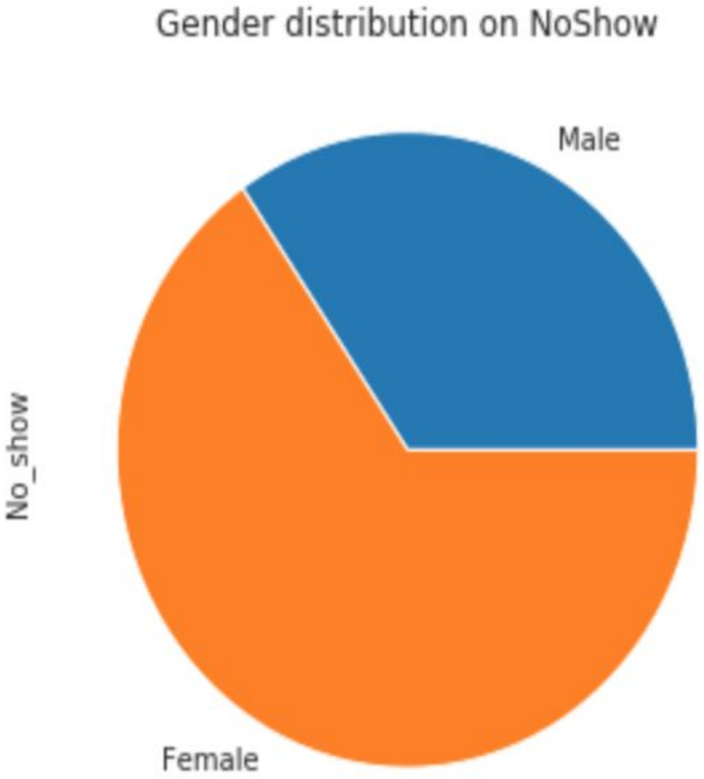
- No duplicate records
- Mix of categorical and numeric features
- Age minimum value = -1 (how?)
- **New variable**: Wait Days = AppointmentDate - ScheduledDate
- Cases with Wait Days < 0 (how?)
- **Discretization**: Age and WaitDays
- **One Hot Encoding**: All categorical variables
- **Final Patients = 110521**

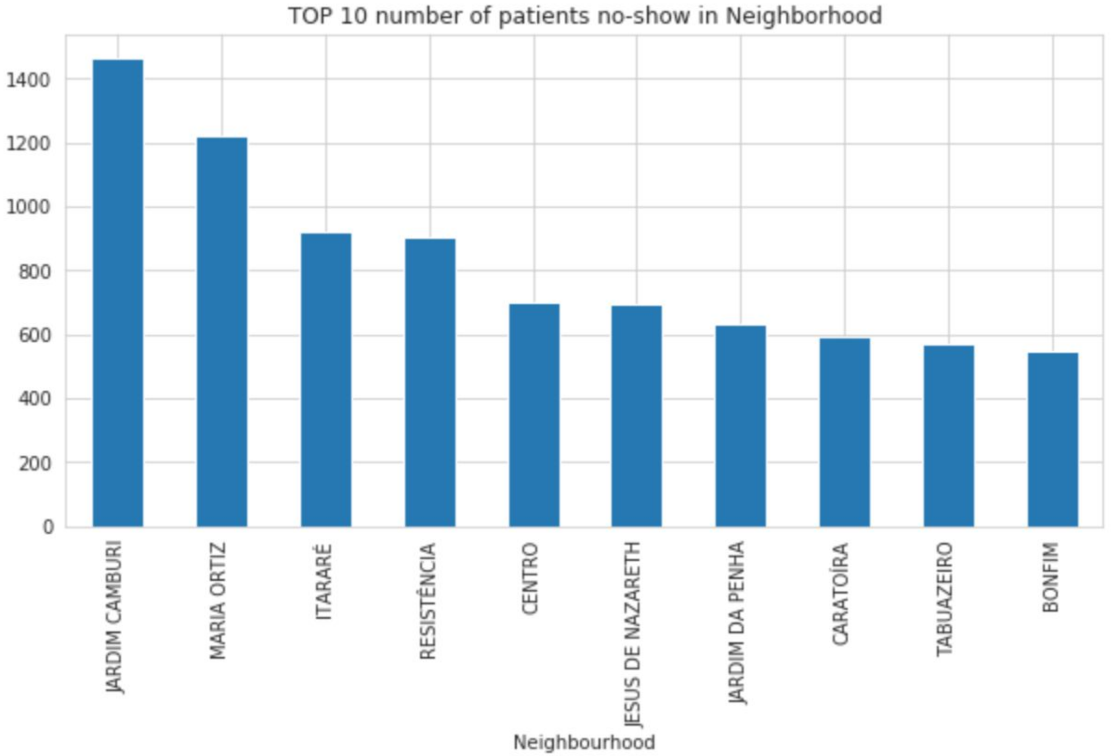
Exploratory Data Analysis

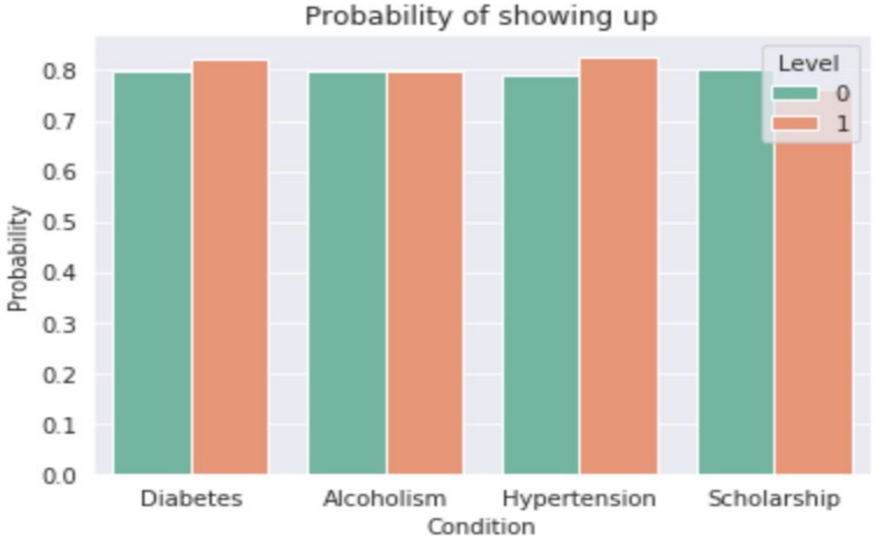


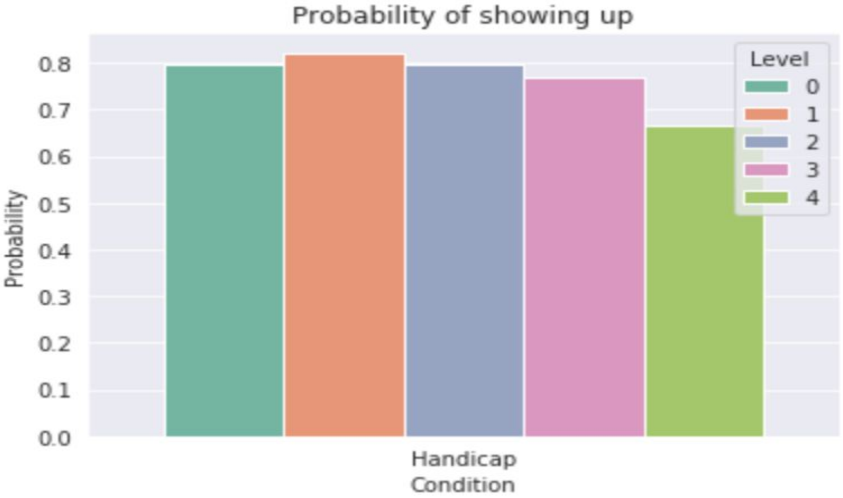


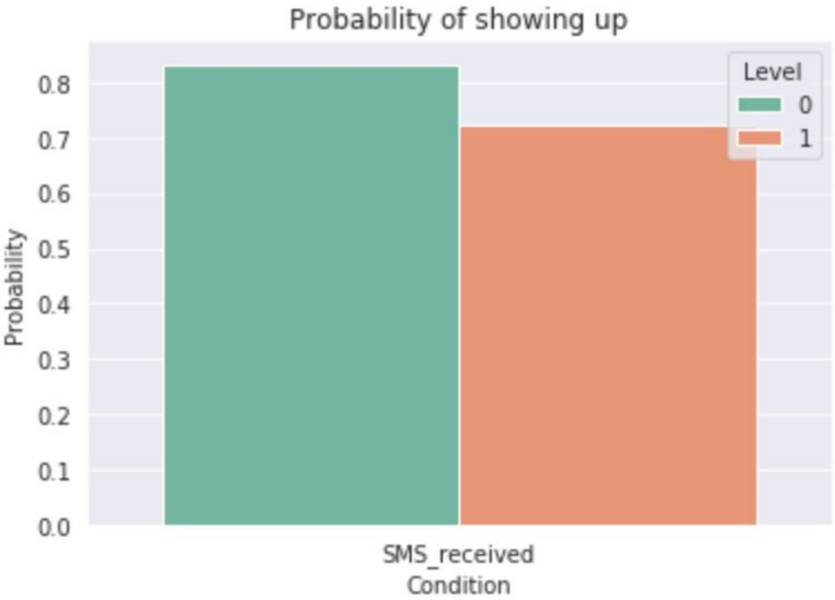
Exploratory Data Analysis

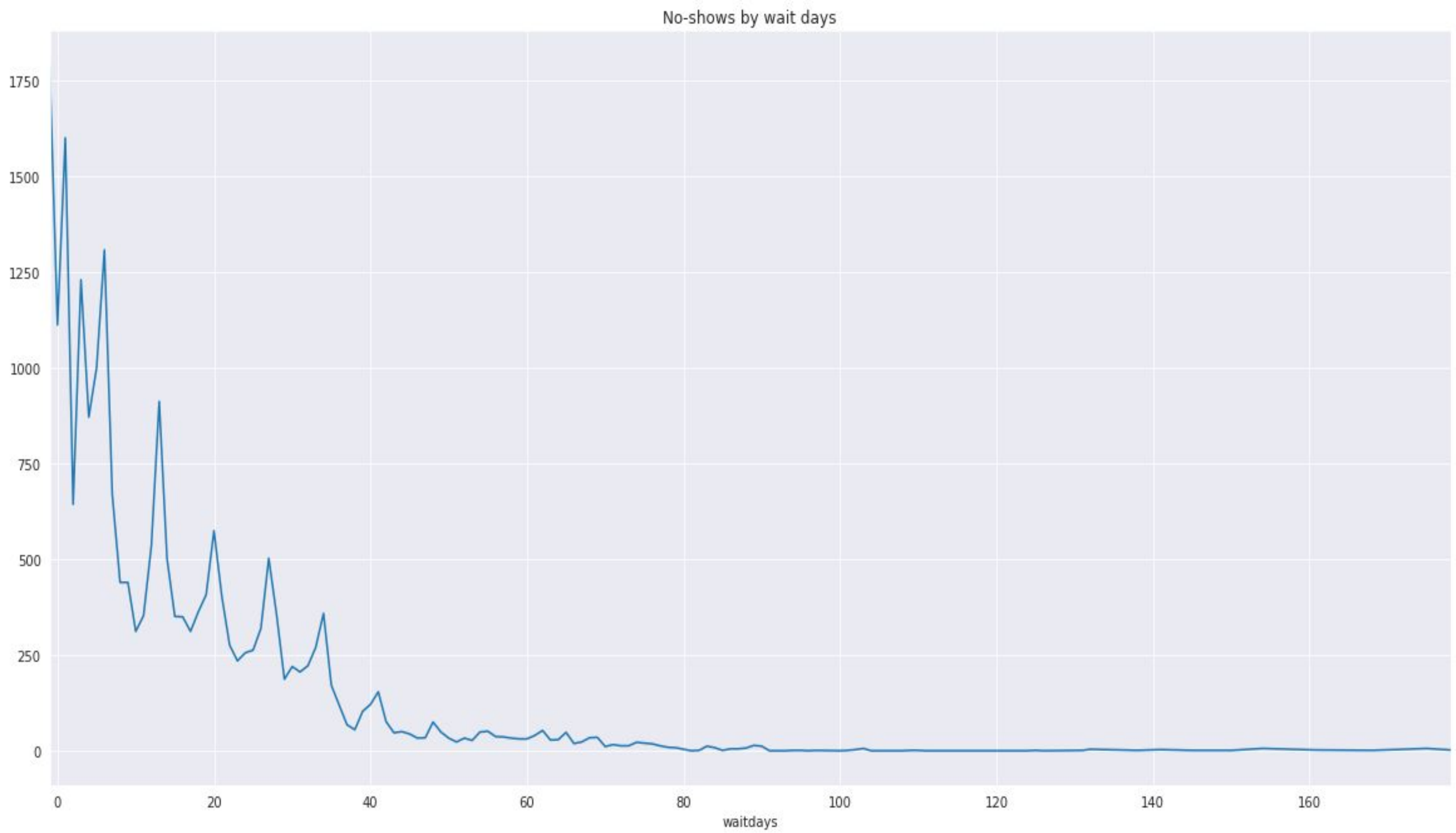





Probability of showing up
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Diabetes
Alcoholism
Hypertension
Scholarship
Condition
Level
0
1


Probability of showing up
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Handicap
Condition
Level
0
1
2
3
4


Probability of showing up
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
SMS_received
Condition
Level
0
1


No-shows by wait days
1750
1500
1250
1000
750
500
250
0
0
20
40
60
80
100
120
140
160
waitdays

Classification

Train (0.75) and test (0.25) split.

5-fold cross validation on training data.

Experiment 1: All features

Models

- Majority Baseline
- Logistic Regression
- Random Forest (with importance)
- Random Forest (with modified threshold)

Experiment 2: Important Features

Models

- Logistic Regression
- Random Forest
- AdaBoost
- Random Forest (with modified threshold)

Evaluation

- Accuracy
- Precision (macro average)
- Recall (macro average)
- F1 score (macro average)
- Area under ROC curve: used for refitting models

Results

Hypothesis: Diabetes, Hypertension, Alcoholism and Handicap do not impact no-shows

Experiment 1: All features

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Majority Baseline	0.798	0.39	0.5	0.438
Logistic Regression	0.795	0.58	0.5	0.46
Random Forest	0.798	**0.74**	**0.5**	0.45
Random Forest (modified threshold)	0.595	**0.62**	**0.68**	0.56

Results

FEATURE	IMPORTANCE	FEATURE	IMPORTANCE
Age	0.13	Handicap	0.01
Wait Days	0.75	Hypertension	0.01
Gender	0.01	Diabetes	0
SMS Received	0.07	Alcoholism	0.01

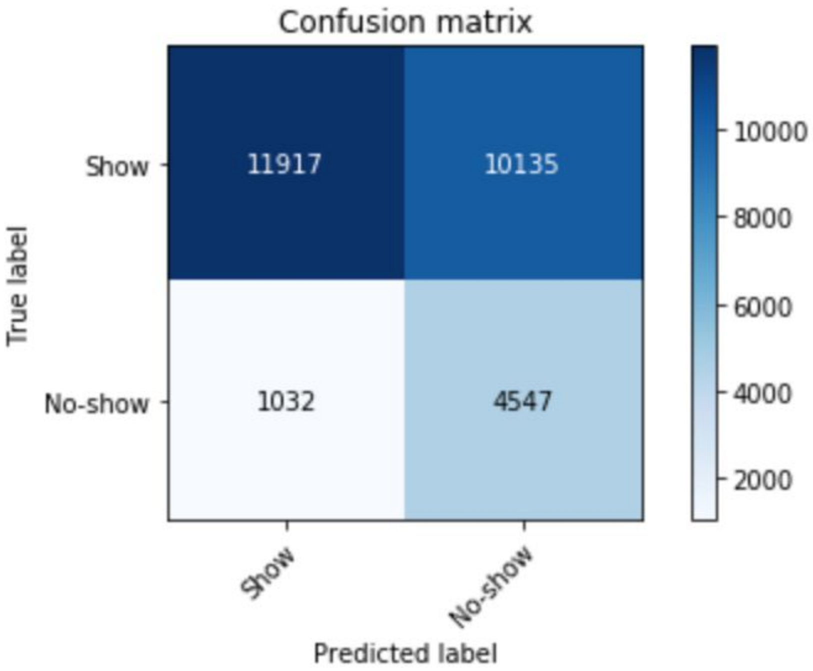
Results

Experiment 2: Age, Wait Days and SMS Received

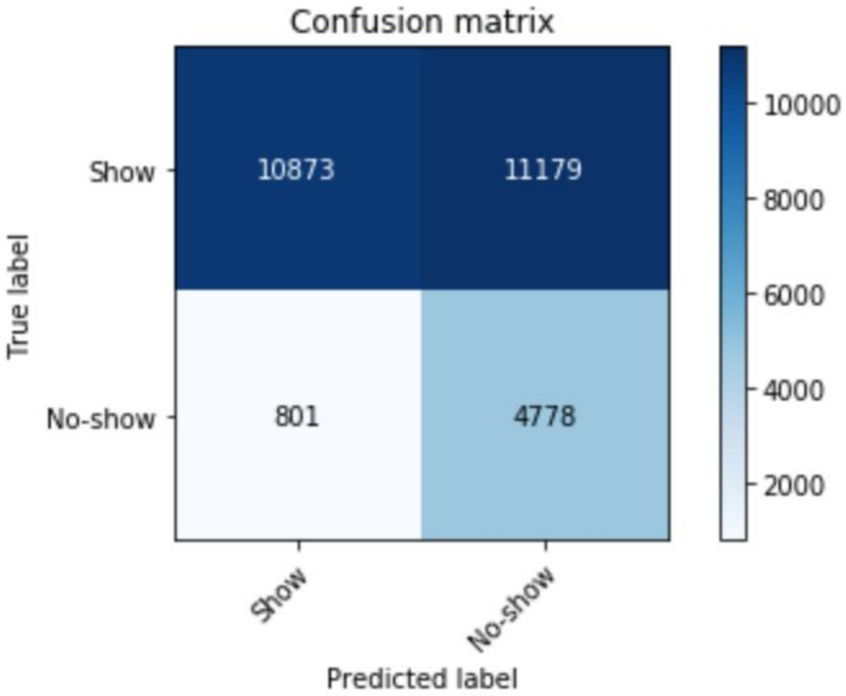
MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Logistic Regression	0.798	0.4	0.5	0.44
Random Forest	0.798	0.4	0.5	0.44
Random Forest (modified threshold)	0.553	0.55	0.55	0.55
AdaBoost (modified threshold)	0.566	**0.62**	**0.67**	0.54

Modified Threshold

RANDOM FOREST



ADABOOST



Conclusion and Future Work

- Best performance: **Random Forest and AdaBoost with modified AUC thresholds**
- Using all features is better in this case than subset of features
- Interventions based on important features
    - Age
    - SMS Received
    - Wait Days
- Future work:
    - Neighborhood (77 unique value)
    - Feature selection techniques

Questions

- Language: Python
- Libraries used: Pandas, scikit-learn, jupyter notebook, matplotlib, seaborn