HOW CAN WE PREDICT "GENDER" BY PERSONAL PREFERENCES AND FEARS IN 90% ACCURACY?

Link: http://rpubs.com/zcy1121/gender

PROJECT 2

GROUP NAME:

ALLA

GROUP MEMBERS:

RUYUE ZHANG,

PHOEBE WU,

LIWEI ZHU,

SHENG LUO

DATASET

- Preferences ⟷ Gender
- Data source: Kaggle, Young people survey
- The dataset includes 9 groups of data:

Music Preferences, Movie Preferences, Hobbies & Interests, Phobias, Health Habits, Personality traits, Views on life & Opinions, Spending habits, Demographics

All values are classified into five levels (1~5) except demographic.

- We select:

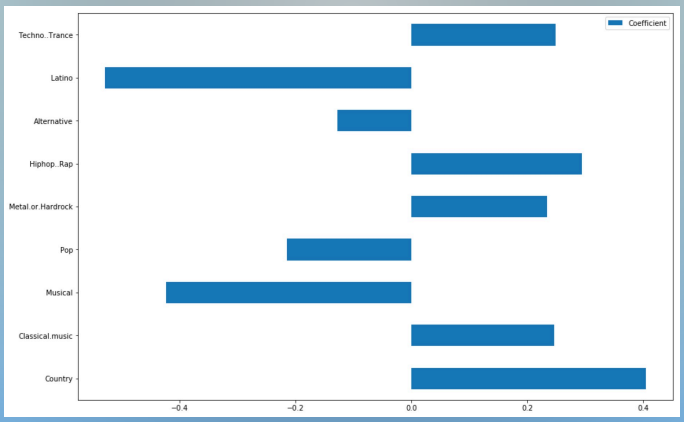
Music Preferences: 19 variables, Movie Preferences: 12 variables, Hobby and Interests: 32 variables, Phobias: 10 variables

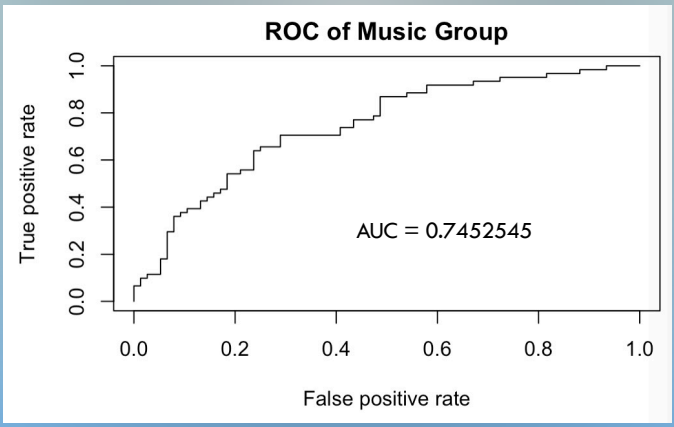
- Two main steps of solution:

1.Build model within groups. (logistic regression)

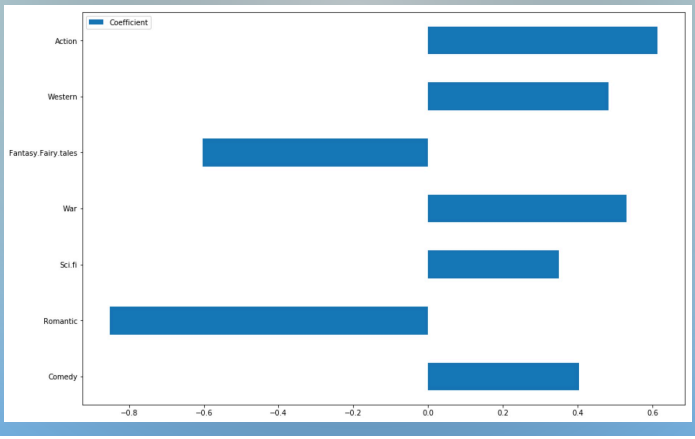
2.Build model in the final group. (logistic regression and KNN)

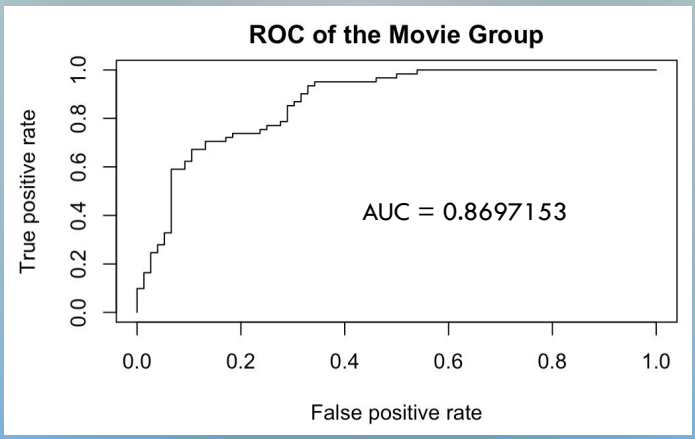
LOGISTIC REGRESSION OF MUSIC GROUP



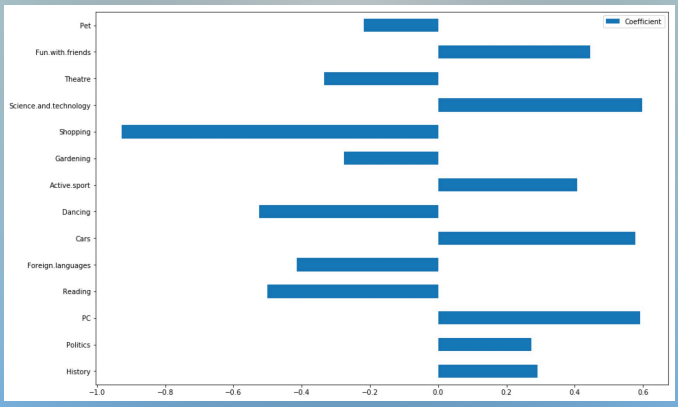


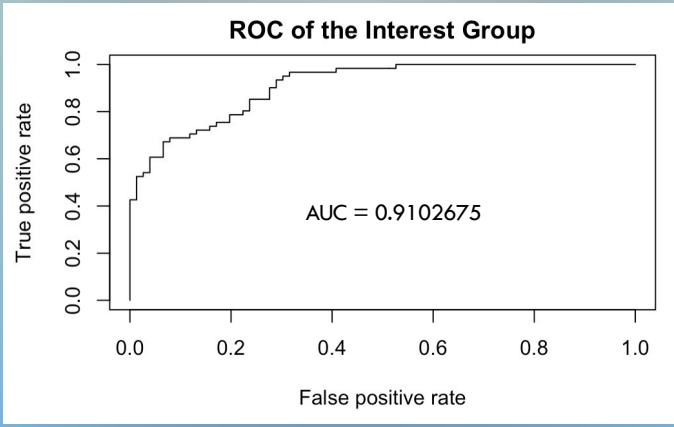
LOGISTIC REGRESSION OF MOVIE GROUP



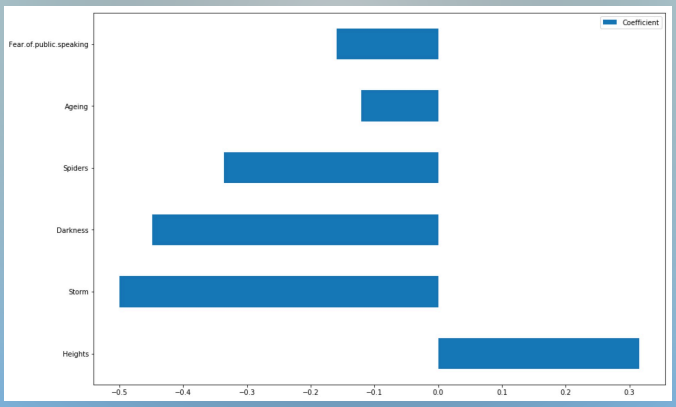


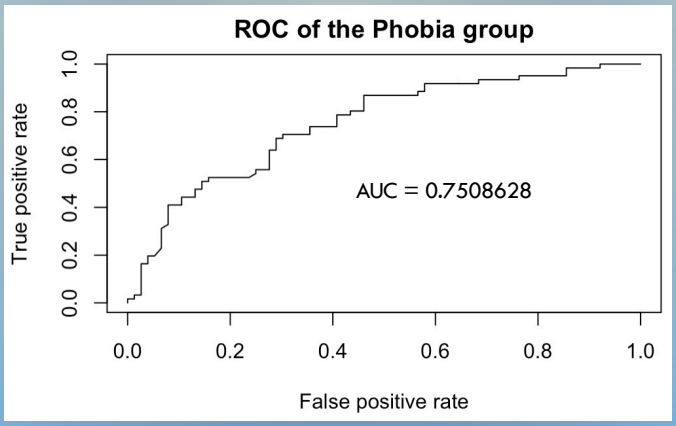
LOGISTIC REGRESSION OF INTEREST GROUP



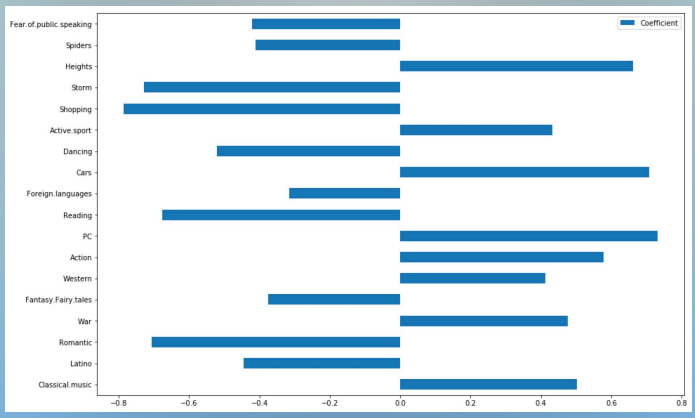


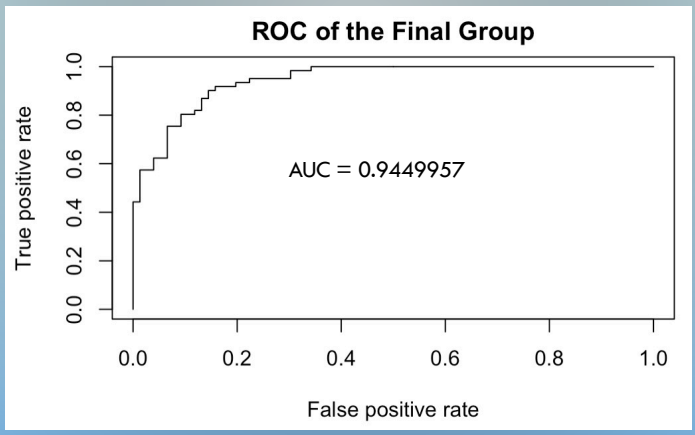
LOGISTIC REGRESSION OF PHOBIA GROUP



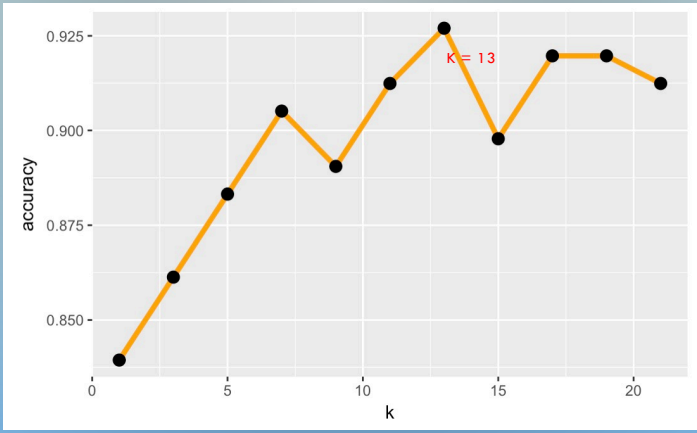


LOGISTIC REGRESSION OF FINAL GROUP





K-NN MODEL OF FINAL GROUP

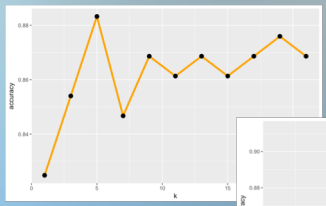


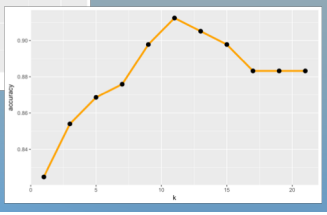
CONCLUSION

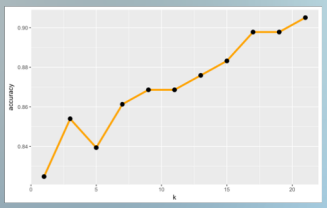
- Based on "Classical music, Latino, Romantic, War, Fantasy fairy tales, western, Action, PC, Reading, Foreign languages, Cars, Dancing, Active sport, Shopping" preferences and fear of "Storm, Heights, Spiders, Public speaking", we can predict gender in 92% chance right.

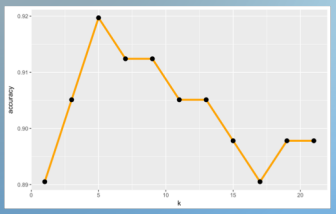
LIMITATIONS

- Sample size is not big
- We lost a lot observations data after omitting missing values.









What can we do next?

- Collect more samples

Thank You