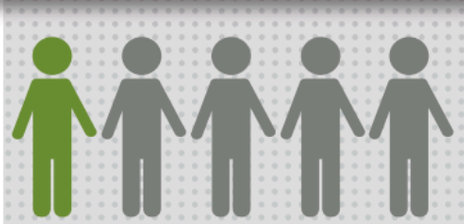
# Predicting Who Will Seek Mental Health Treatment in the Tech Workplace

*Kevin Hahn*

# May is Mental Health Month

## Mental Health Facts
### IN AMERICA

**Fact:** 43.8 million adults experience mental illness in a given year.



1 in 5 adults in America experience a mental illness.

Nearly 1 in 25 (10 million) adults in America live with a serious mental illness.

One-half of all chronic mental illness begins by the age of 14; three-quarters by the age of 24.

*https://www.nami.org/NAMI/media/NAMI-Media/Infographics/GeneralMHFacts.pdf*

# Context

- Prevalence
  - National Alliance on Mental Illness (NAMI) and National Institute of Mental Health (NIMH) reported that 60% of adults with mental illness received no treatment services in the past year
- In 2015, estimated 43.4 million (17.9% of U.S. adults) had a mental illness (NIMH, 2015)

# Context

- Impact
  - CDC (2017): From 1981 to 2015, suicide is tenth leading cause of death for all adults
  - Insel (2008): In 2002 alone, Serious Mental Illness cost $317.6 billion in disability benefits, healthcare expenditures, and lost earnings

# Context

- Impact in the workplace (Mental Health America, 2017)
  - Depression as most common and costly disorder
  - Employees may not seek out treatment due to fear of effect on job

# Dataset

- Open Sourcing Mental Illness (OSMI) 2014 Mental Health in Tech Survey
  - Believed to be largest survey on mental health in tech industry at that time
  - Measures “attitudes toward mental health and frequency of mental health disorders in the tech workplace”

# Research Question

What factors predict whether tech employees will seek mental health services?

# Variables

| | Timestamp | Age | Gender | Country | State | Self-Employed | Family History | Work Interfere | Treatment |
|---|---|---|---|---|---|---|---|---|---|
| Data Type | Timestamp | Integer | String | String | String | Binary | Binary | Binary | Binary |
| Has NaNs? | | | | | Yes | Yes | | | |
| Cleaned? | | Yes | No* | | | | | No* | |
| *N* | 1259 | 1259 | 1259 | 1259 | 744 | 1241 | 1259 | 995 | 1259 |

# Variables

| | Number Employees | Remote Work | Tech Company | Benefits | Care Options | Wellness Program | Seek Help | Anonymity | Leave |
|---|---|---|---|---|---|---|---|---|---|
| Data Type | Categorical | Binary | Binary | Categorical | Categorical | Categorical | Categorical | Categorical | Categorical |
| *N* | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 |

| | MH Consequences | PH Consequences | Coworkers | Supervisor | MH Interview | PH Interview | Mental vs. Physical | Observed Consequences | Comments |
|---|---|---|---|---|---|---|---|---|---|
| Data Type | Categorical | Categorical | Categorical | Categorical | Categorical | Categorical | Categorical | Binary | String |
| Has NaNs? | | | | | | | | | Yes |
| *N* | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 1259 | 164 |

# Initial Model

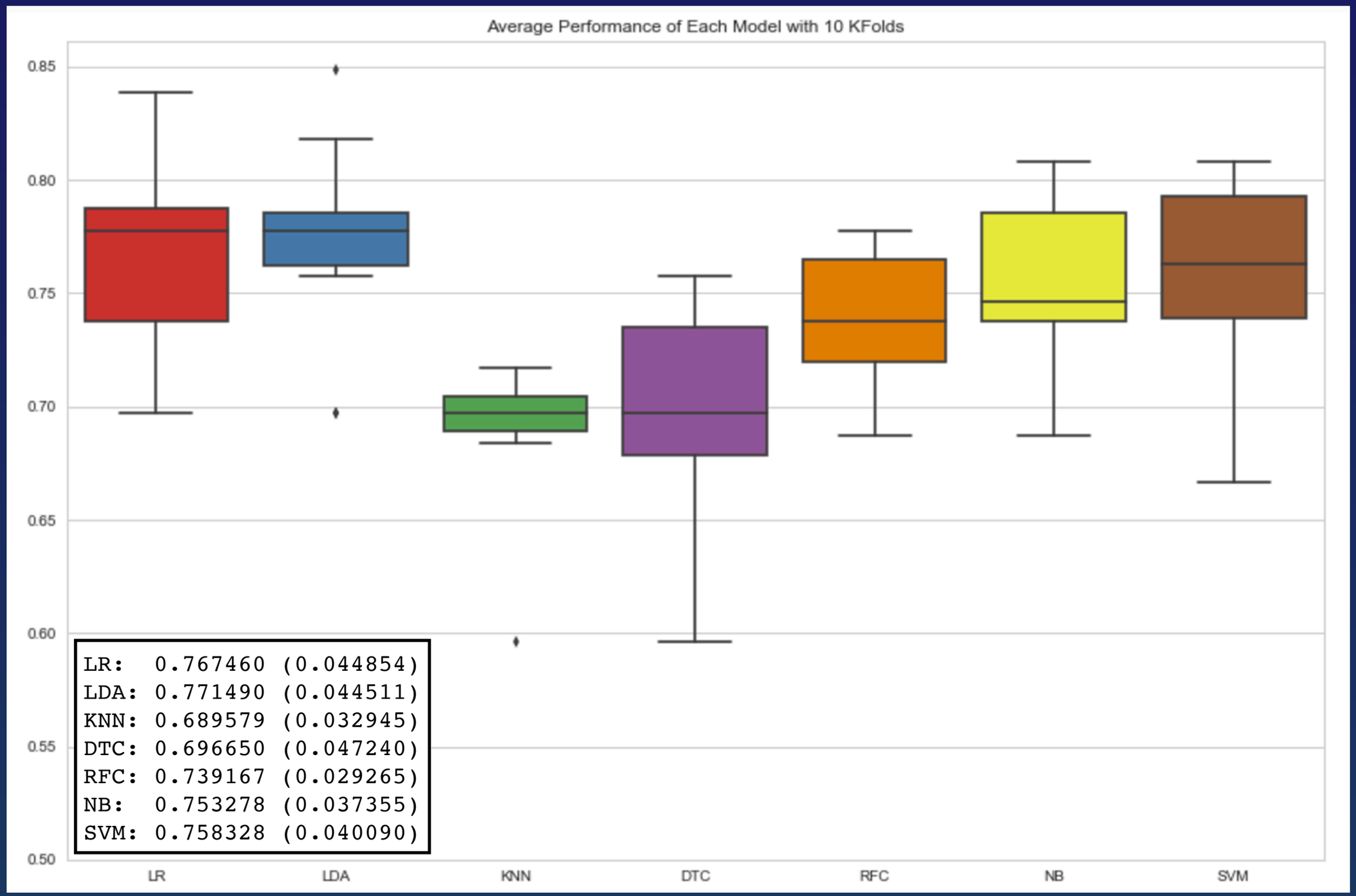
- *N* = 989
  - Dropped observations with ages below 18 and over 200
  - Dropped observations containing NaNs (work_interfere)
- Features = 22
  - Dropped features with highly imbalanced classes (Gender, Country, self_employed)
  - Dropped features with high frequency of missing data (US State, Comments)
  - Dropped Timestamp
  - Added 2 engineered features (US/non-US, Age < 40)
- Outcome = Have you sought treatment for a mental health condition?  (yes/no)

# Initial Model

- *N* observations = 989, *p* predictors = 22
- Applied seven classifier methods:
  - Logistic Regression (LR)
  - Linear Discriminant Analysis (LDA)
  - K-Nearest Neighbor Classifier (KNN)
  - Decision Tree Classifier  (DTC)
  - Random Forest Classifier (RFC)
  - Naive Bayes (NB)
  - Support Vector Machines Classifier (SVM)
- Used 10 k-folds cross-validation with mean accuracy (standard deviation accuracy)

```
LR:  0.767460 (0.044854)
LDA: 0.771490 (0.044511)
KNN: 0.689579 (0.032945)
DTC: 0.696650 (0.047240)
RFC: 0.739167 (0.029265)
NB:  0.753278 (0.037355)
SVM: 0.758328 (0.040090)
```

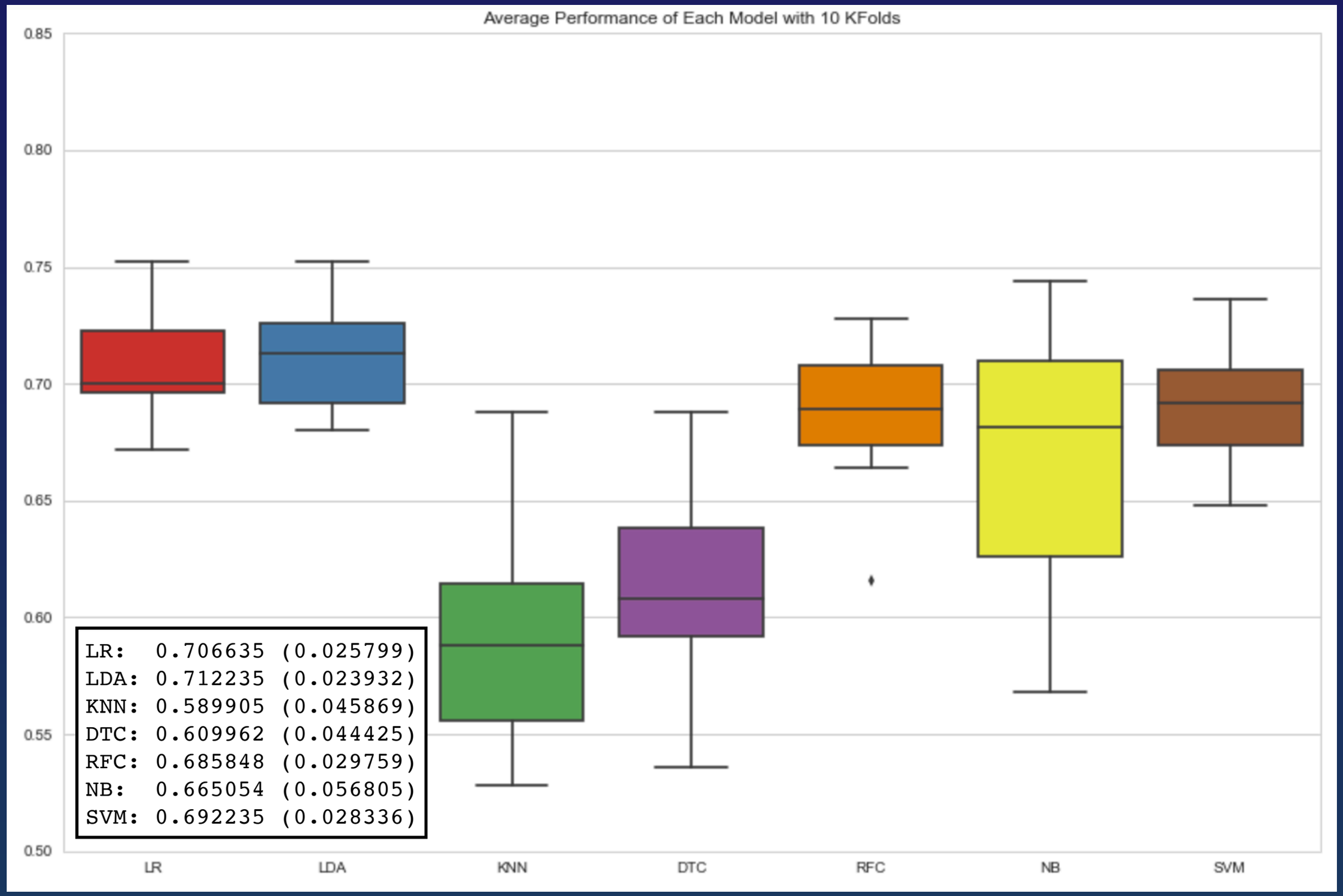
# Initial Model



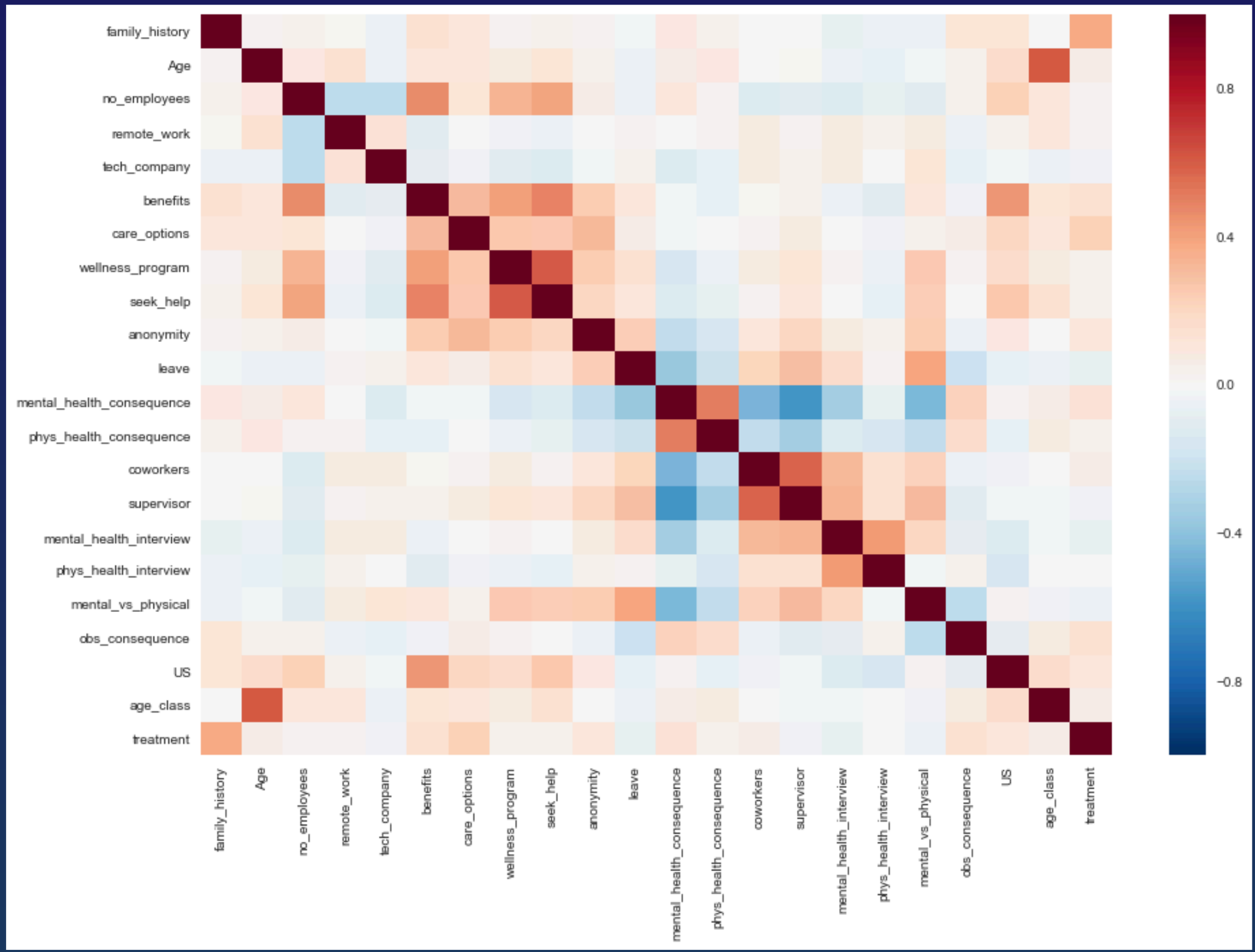
# Revised Initial Model

- *N* = 1251
  - Removed work_interfere feature which increased *N*
- Features = 21
  - Dropped work_interfere feature (264 N/As):
    - If you have a mental health condition, do you feel that it interferes with your work? (Never, Rarely, Sometimes, Often)
    - Item is confusing and potentially misleading
    - Another way of stating outcome variable, or confirming participant mental illness?
- Outcome = Have you sought treatment for a mental health condition?  (yes/no)

# Revised Initial Model



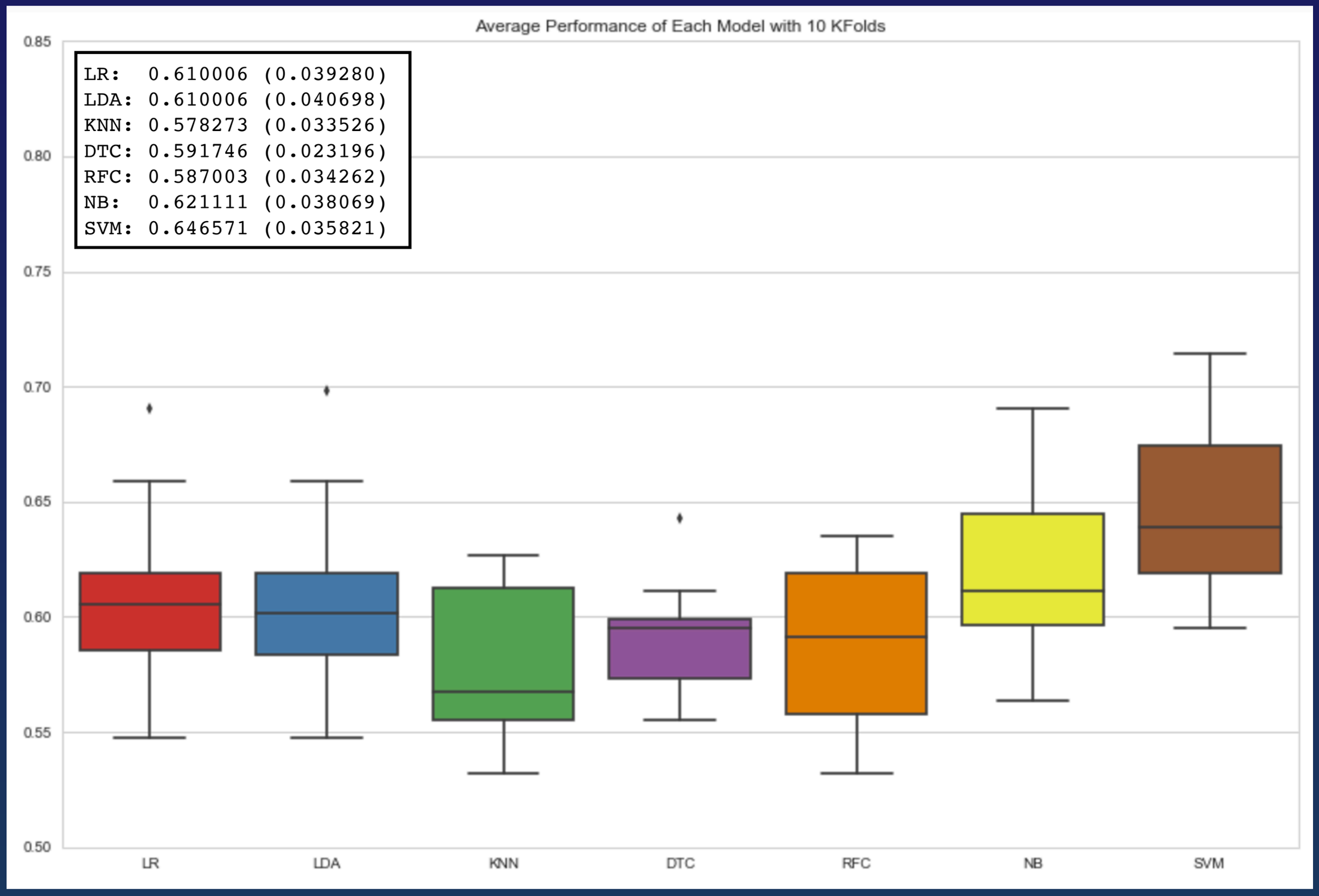
# Feature Reduction



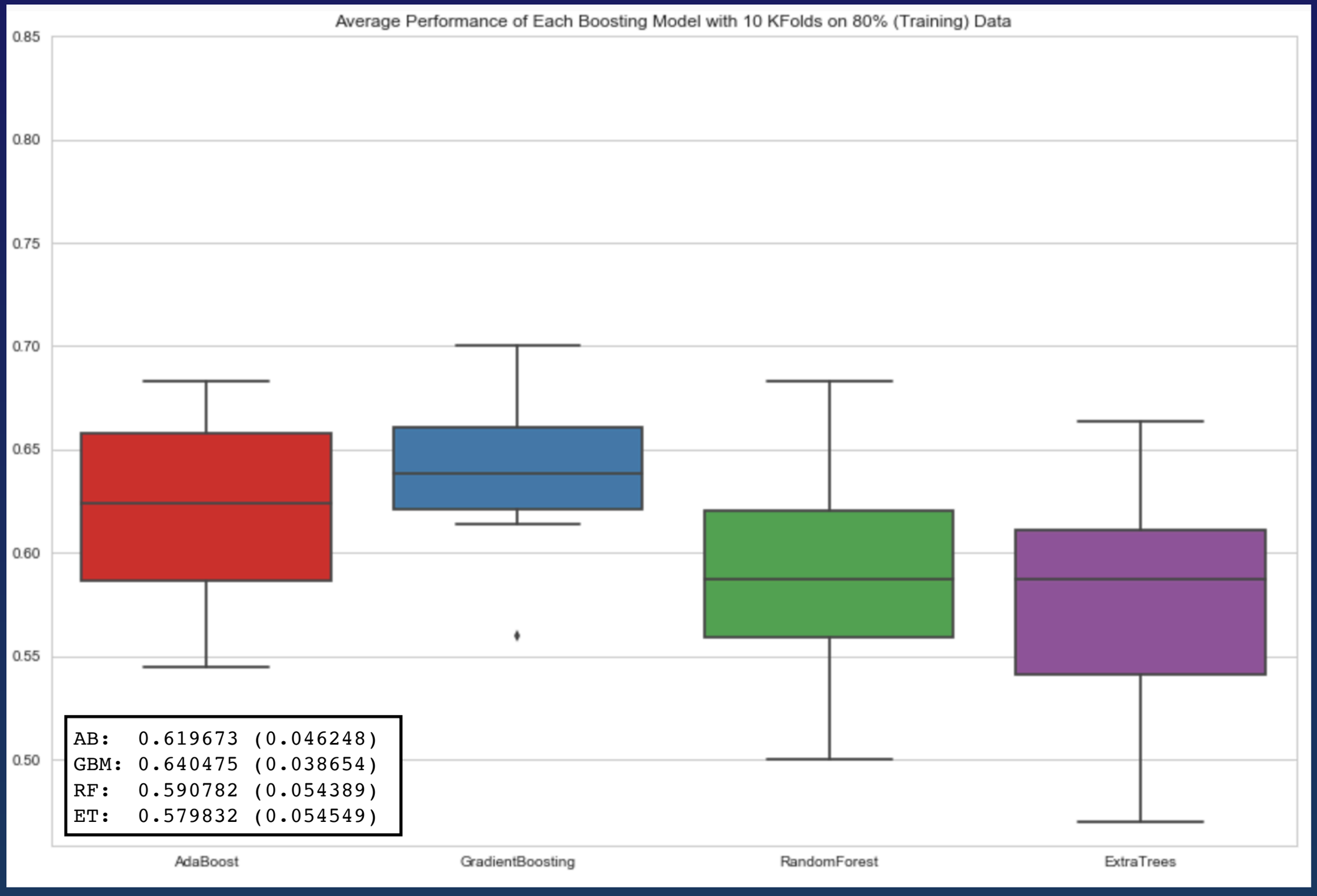
# Reduced-Features Model

- *N* = 1259
  - Maximum N: Used only features that all participants answered
- Features = 6
  - (1) Number of employees, (2) ease of taking MH leave, (3) knowledge about available MH care options, (4) MH benefits, (5) availability of workplace wellness program, (6) employer-provided resources to seek help
  - Includes generally "actionable" features (e.g., employee education, workplace factors)
- Outcome = Have you sought treatment for a mental health condition?

# Reduced-Features Model



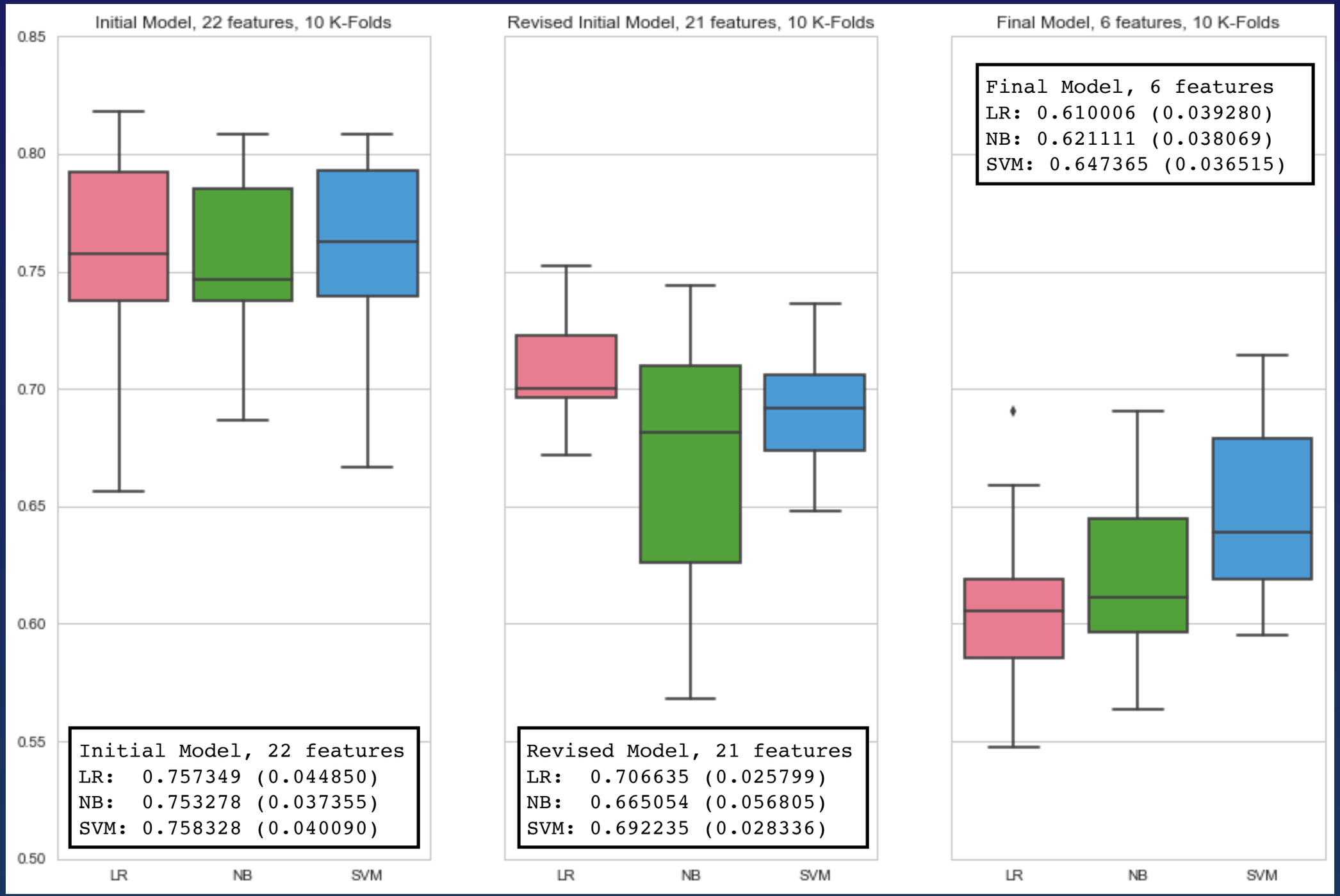
# Boosting Models



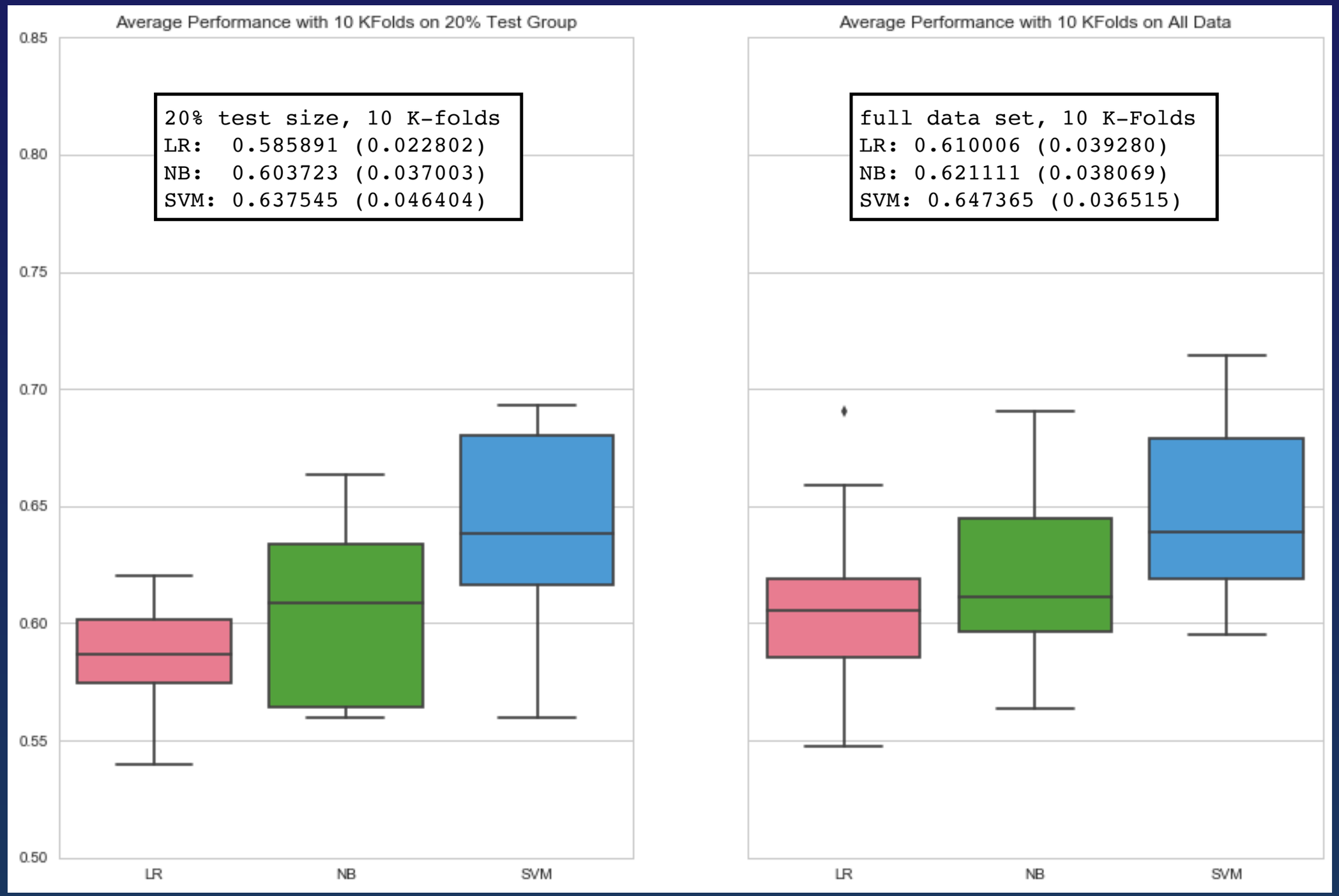
# Hyperparameter Tuning

- Tuned select classifiers in reduced-features model:
  - Logistic Regression
  - Random Forest
  - SVM Classifier
- Minimal increases in performance
- Kept Logit, Naive Bayes, and SVM Classifier for final model

# Comparison of Initial, Revised, and Reduced-Features Models



# Test Group vs. Full Dataset



# Outcome

The size of the organization, ease of taking leave, employees' knowledge about available mental health care options, availability of mental health benefits, availability of a workplace wellness program, and employee-provided resources about mental health resources may predict which tech employees will seek mental health services with approximately 60-65% accuracy.

# Methodological Concerns

- Directionality/Causality
  - Does seeking mental health services predict having workplace benefits? Vice versa?
- Survey Items
  - Comments from participants about needing N/A or I don't Know for certain items (e.g., family history)
  - Words used interchangeably: Condition, Diagnosis, Illness

# Extensions

- Compare to OSMI's 2016 dataset (data collection still in progress!)
  - Validate model against similar outcome variable
    - Correlate responses using Likert-type scales
- Perform Text Analysis
  - Employ other text analyses (e.g., n-grams, NLP)
  - Determine how text supports or fails to support model

# Conclusion

- Providing employee education about mental health care options and offering a workplace wellness program may predict tech employees' behaviors to seek mental health treatment
- Other factors such as the size of the organization and availability of mental health benefits may also predict which tech employees will seek mental health treatment
- These findings may or may not be representative for employees in other organizations

# Questions?

Predicting Who Will Seek Mental Health Treatment in the Tech Workplace

https://github.com/KevinHahnMTBC/Mental-Health-and-Tech