

Understanding factors influencing loneliness and corresponding Quality of Life via interviews, smartphone sensing and data modeling

Thesis defense of MSc in Computer Science by Adam Honoré

Supervised by Katarzyna Wac

Presented 22/08/2017



Presentation Outline

- ❖ Topic Introduction
- ❖ Research Question
- ❖ Methodology
- ❖ Data Analysis
- ❖ Research Findings
- ❖ Conclusive Remarks
- ❖ Future Work

Topic Introduction - Loneliness and QoL

“Loneliness is the unpleasant experience that occurs when a person's network of social relations is deficient in some important way, either quantitatively or qualitatively.” - Perlman and Peplau

“WHO defines Quality of Life as an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.” - WHOQOL Group

- A large amount of people regularly feels lonely
- This is predicted to increase even more in the future
- Loneliness can lead to stress, depression and other psychological effects
- Loneliness may increase the likelihood of early mortality
- Extensive use of internet and social media linked to increased feelings of loneliness



Research Question

Challenge: People don't like to talk about loneliness, about being lonely or the extend of their loneliness

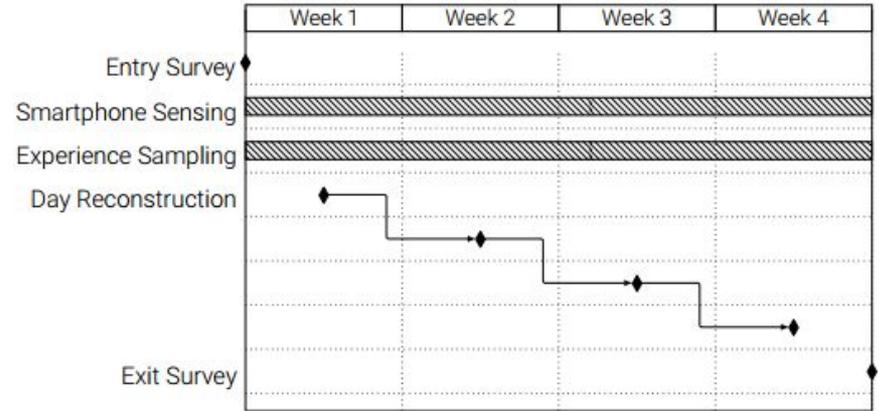
Can we leverage the ubiquity of smartphones in people's lives to help assess this problem?

Methodology

- Introduction

- Recruitment through online questionnaire
- Three subjects participated in study
- Study lasted one month (April - May 2017)

- Mixed-methods strategy
 - Provides depth and breadth



Methodology

- Assessing Loneliness

- UCLA Loneliness Scale
 - Version 1 (1978)
 - Version 2 (1980)
 - Version 3 (1996)
- Three-Item Loneliness Scale (2004)
 - Short version of UCLA Loneliness Scale for large-scale surveys
- Danish Loneliness Scale (2007)
 - Translation of UCLA Loneliness Scale

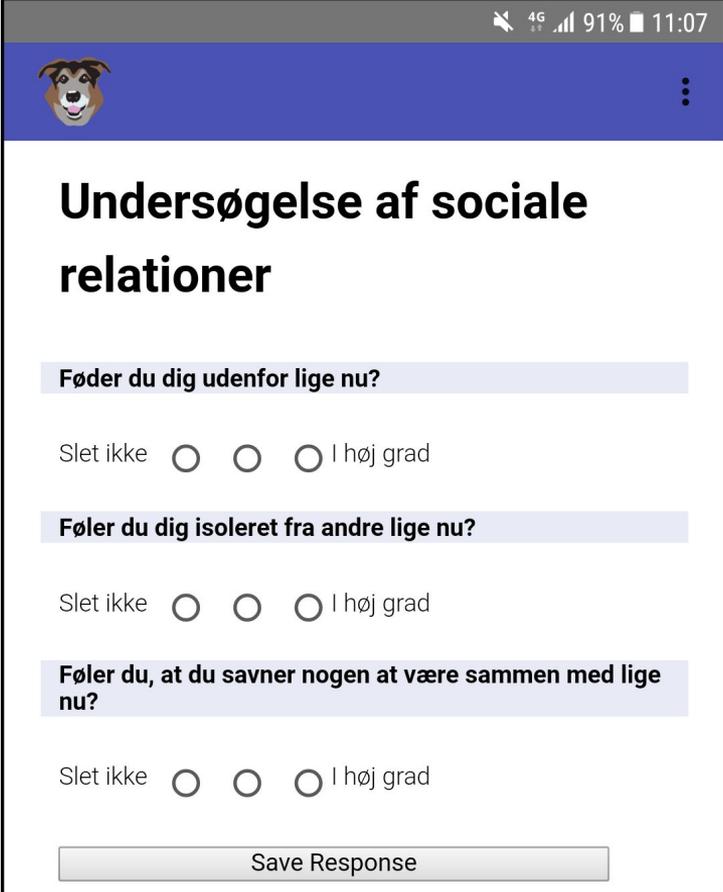


Methodology

- Experience Sampling Method

- Flexible for different usage scenarios
- Samples during actual experience
- Prevents recall bias

“a research procedure for studying what people do, feel, and think during their daily lives, It consists in asking individuals to provide systematic self-reports at random occasions during the waking hours of a normal week.” - Larson and Csikszentmihalyi



The screenshot shows a mobile application interface for a survey titled "Undersøgelse af sociale relationer". The app has a blue header with a dog icon and a menu icon. The survey consists of three questions, each with a light blue header and three radio button options: "Slet ikke", a middle radio button, and "I høj grad".

Undersøgelse af sociale relationer

Føder du dig udenfor lige nu?

Slet ikke I høj grad

Føler du dig isoleret fra andre lige nu?

Slet ikke I høj grad

Føler du, at du savner nogen at være sammen med lige nu?

Slet ikke I høj grad

Save Response

Methodology

- Smartphone Sensing: mQoL-logger

- Developed by mQoL Living Lab
- Utilizes ubiquity of smartphones
- Collects quantitative data from a range of smartphone sensors both periodically and on specific events
- Automatically synchronizes data with remote mQoL server
- Anonymizes data sent to server



Methodology

- Day Reconstruction Method

- Assess how subjects' spend their time in the last 24 hours
- Complete overview of the whole day
- Doesn't disturb actual experiences
- Evokes context to ensure good recall

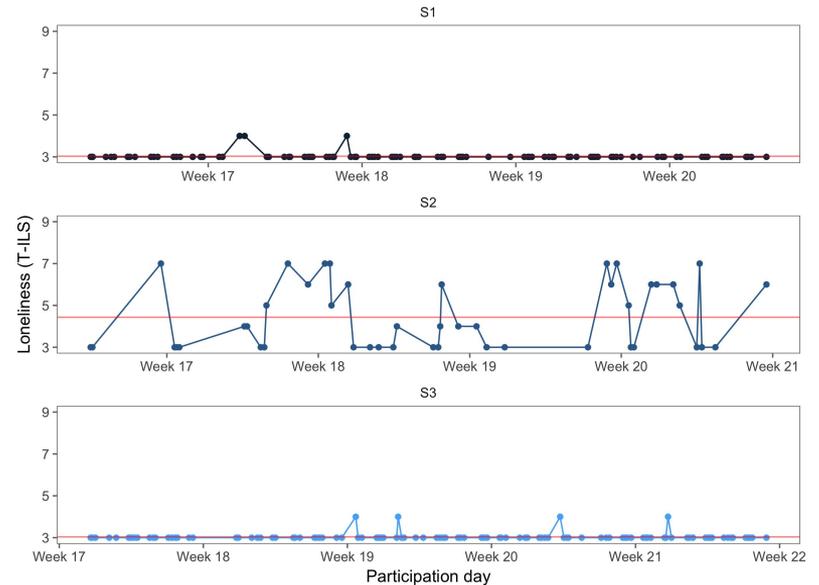
- Used with smartphone sensing data to provide additional context as part of mixed-methods approach

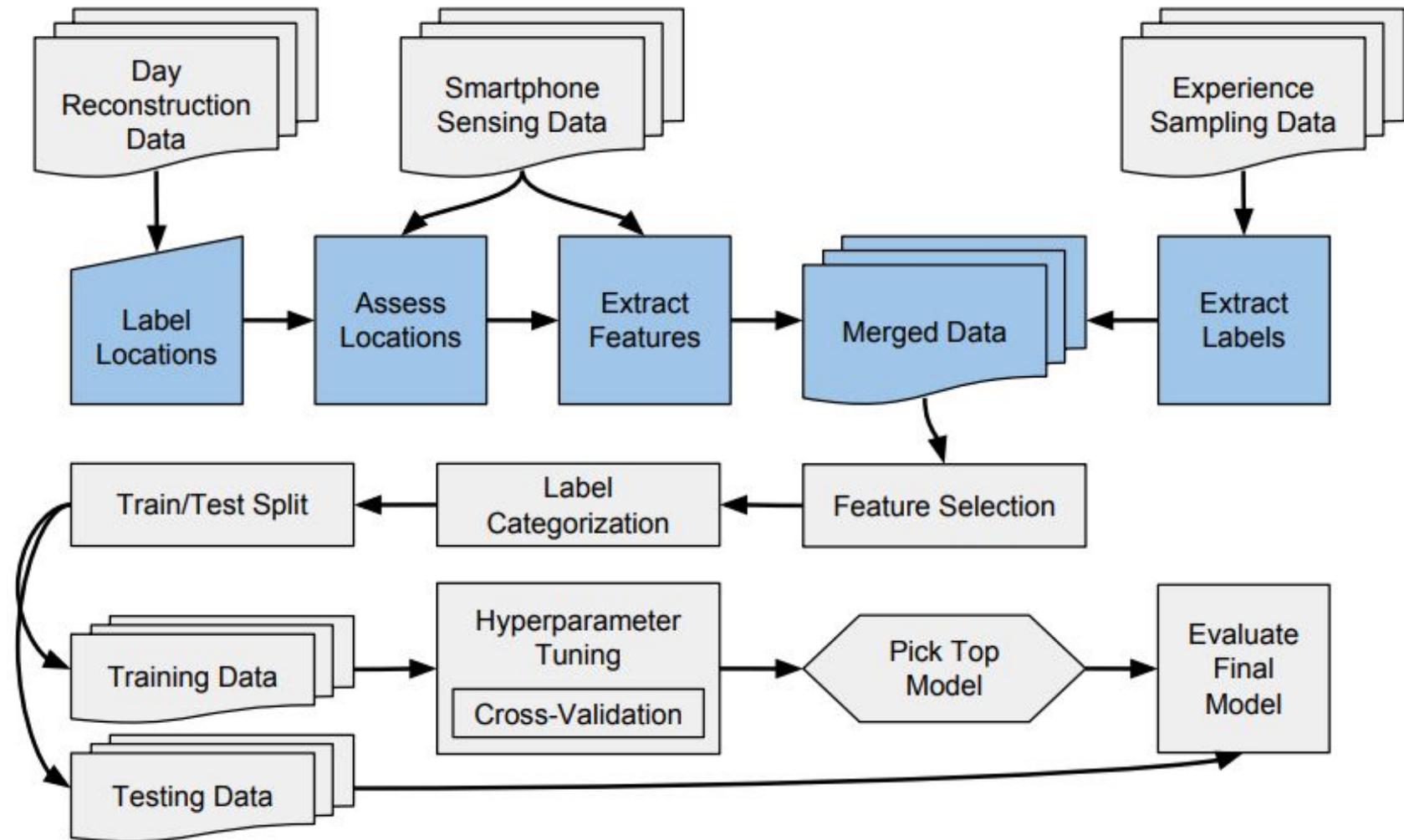


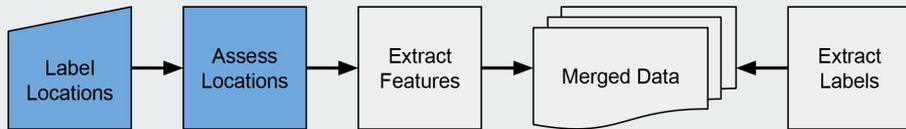
Collected Data Summary

Subject	Expected Responses	Collected Responses	Coverage
S1	120	97	80.83%
S2	120	46	38.33%
S3	120	106	88.33%

Subject	Expected Hours	Collected Hours	Coverage
1	720	393.95	54.72%
2	720	342.67	47.59%
3	720	331.90	46.10%





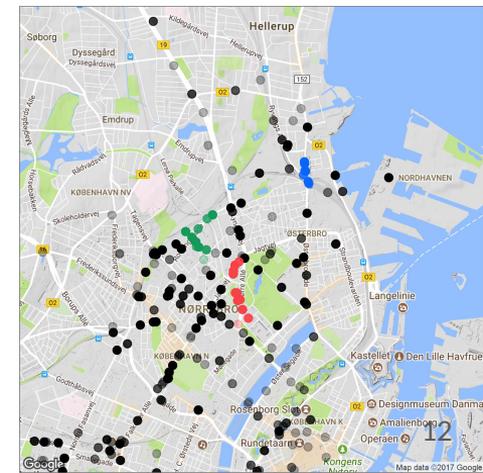
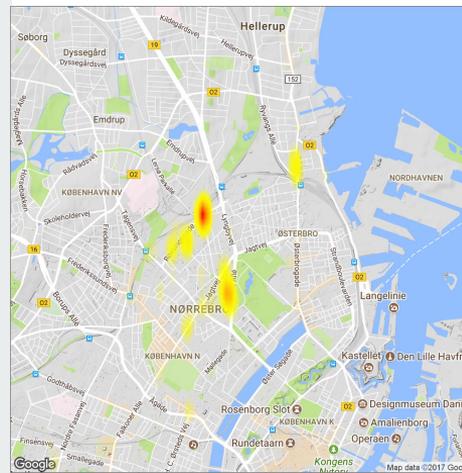


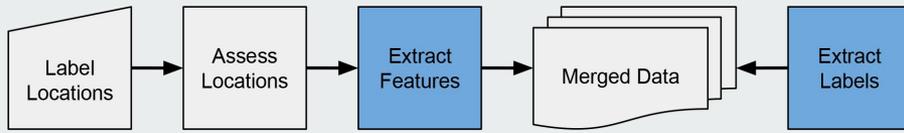
Data Analysis (I)

- Location Assessment

- Only Cell IDs available - no GPS
- OpenCellID cell tower database
- DBSCAN data clustering
- Labels from Day Reconstruction

Cluster	Count	Percentage
Home	9826	52.83%
School	4695	25.24%
Work	844	4.54%
None	3233	17.38%

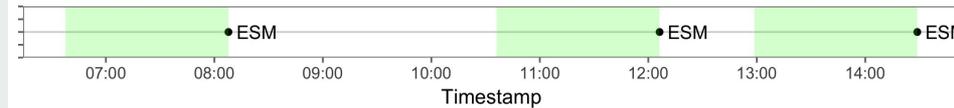


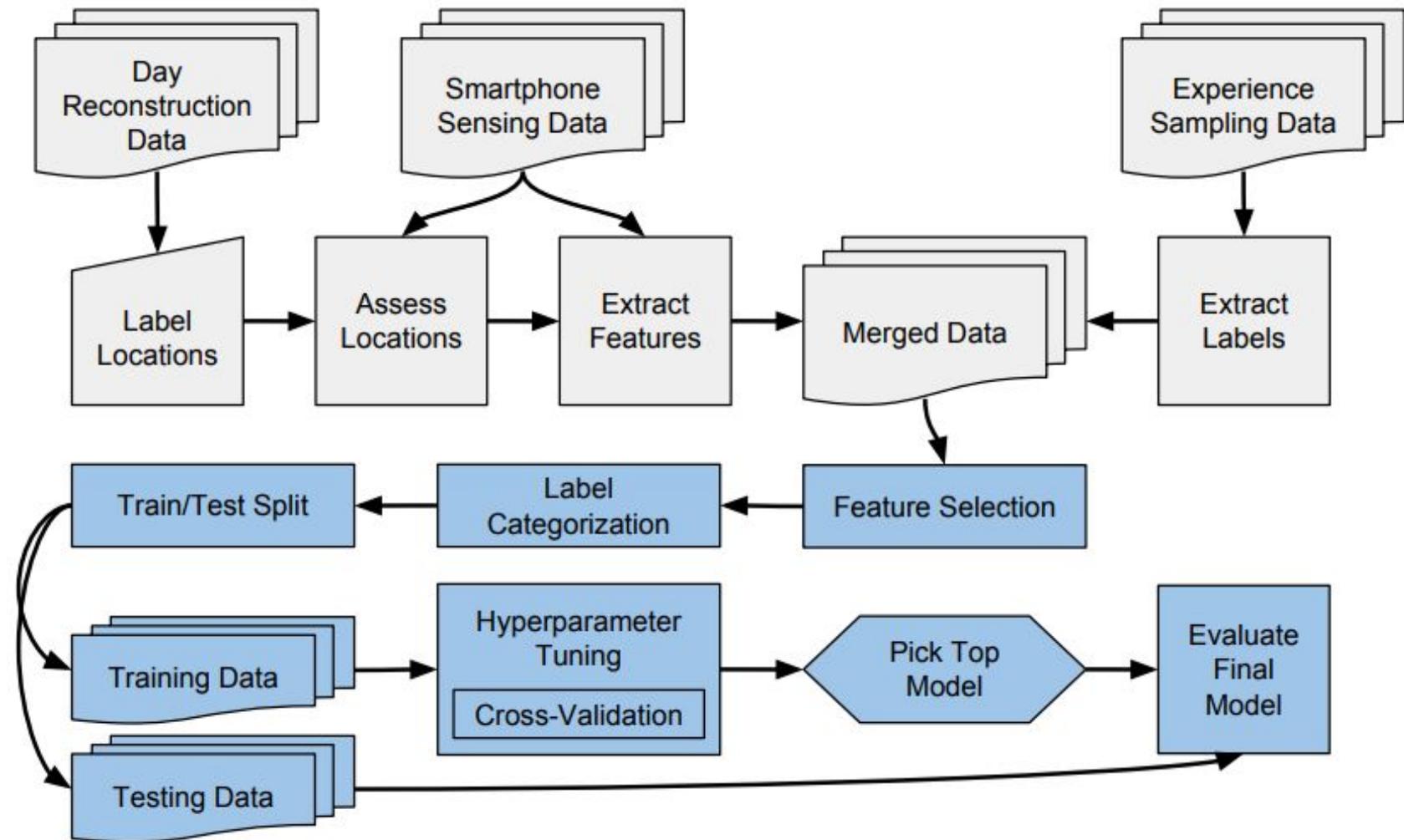


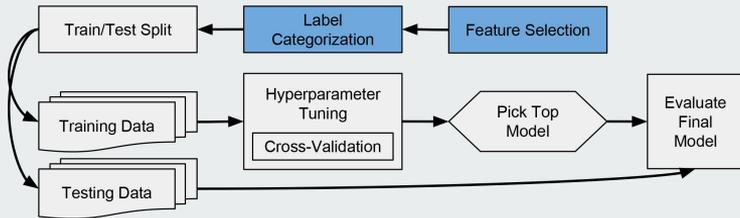
Data Analysis (I)

- Data Merging

- Feature Extraction
 - Summary of data in time window
- Usage counts of common apps
- Time spent in semantic locations
- Time spent connected to networks
- Count of times spent doing specific physical activity (still, walking, running, etc.)
- **46 observations of 47 features**





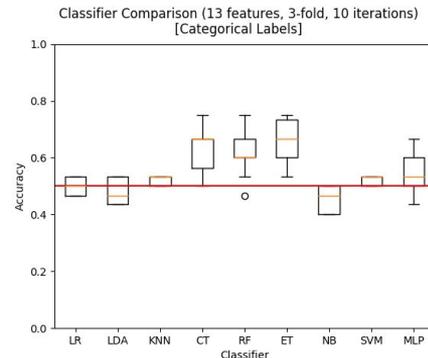
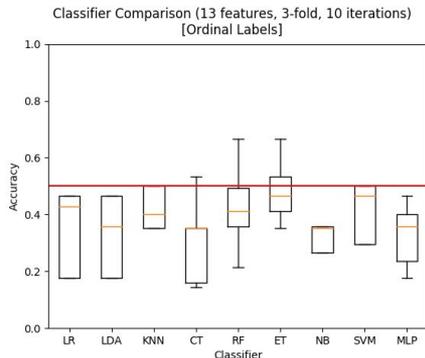


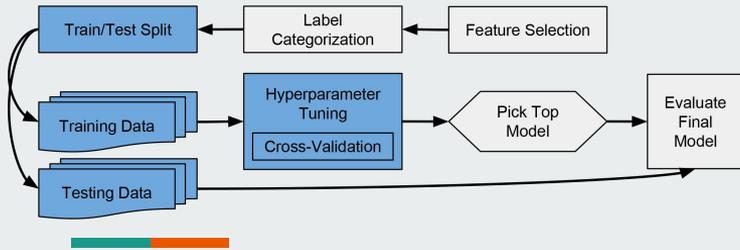
Data Analysis (II)

- Feature Selection & Label Categorization

- Improve computational performance and decrease complexity of models
- Variance threshold
 - Remove features with zero variance
- Variable ranking
 - Select top-13 using ANOVA F-test

Feature	Description
applications_used.calendar_count	Times the Calendar app was opened
applications_used.email_count	Times the Email app was opened
applications_used.messenger_count	Times the Messenger app was opened
applications_used.mobilepay_count	Times the MobilePay app was opened
applications_used.weshare_count	Times the WeShare app was opened
cell_ids_service.cluster_school_percentage	Time spent at school location
cell_ids_service.cluster_unclassified_percentage	Time spent at unclassified location
ping_service.cellular_percentage	Time spent connected to cell towers
ping_service.wifi_home_percentage	Time spent connected to home WiFi
ping_service.wifi_percentage	Time spent connected to WiFi
user_activity.still_percentage	Time spent with phone being still
user_activity.tilting_count	Times phone has been tilting (moving)
user_activity.tilting_percentage	Time spent with phone tilting (moving)

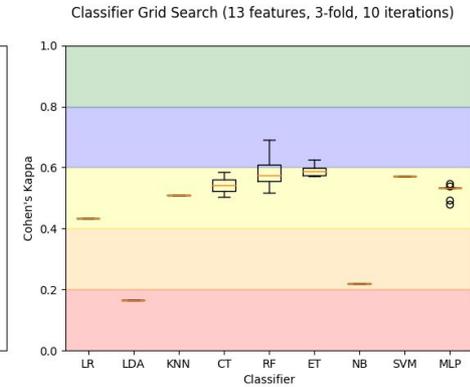
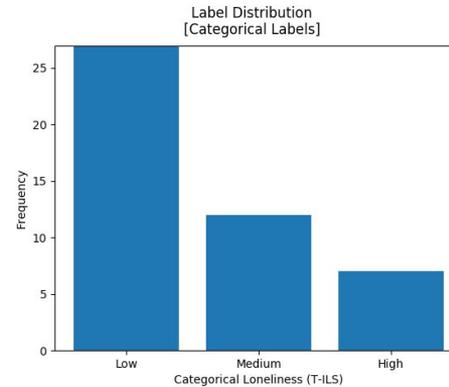


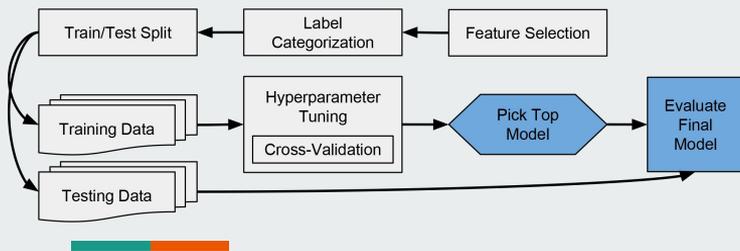


Data Analysis (II)

- Classification

- Data is split into training and testing to reduce overfitting or underfitting
- Nine common classifiers compared
- Cohen's Kappa performance measure
 - Helps with unbalanced classes
- Hyperparameter tuning
- Cross-validation
 - 10 iterations of 3-fold cross-validation

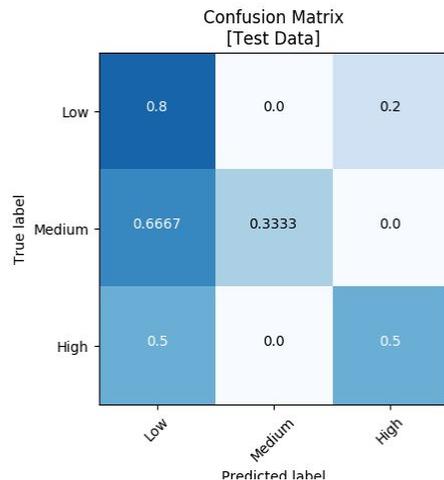
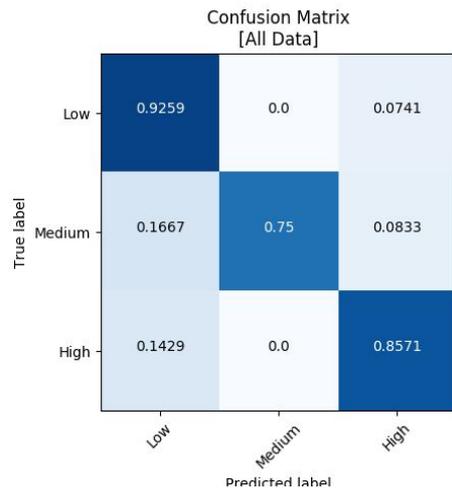




Data Analysis (II)

- Final Model: Random Forest

- Good performance on all data
- Bad performance on test data
- Might be overfitted due to low amount of data rows available



Dataset	κ -performance
Training	0.69
Testing	0.60
Complete	0.87



Research Findings

- Links between physical activity and loneliness, as seen in other studies
- Links between semantic location near other people and loneliness
- No links between social media use and loneliness, as seen in other studies
- **Findings not generalizable**

Rank	Feature	Gini Importance
1	applications_used.email_count	0.23
2	user_activity.still_percentage	0.19
3	cell_ids_service.cluster_school_percentage	0.11
4	ping_service.cellular_percentage	0.09
5	ping_service.wifi_percentage	0.08
6	cell_ids_service.cluster_unclassified_percentage	0.07
7	applications_used.messenger_count	0.07
8	user_activity.tilting_percentage	0.06
9	user_activity.tilting_count	0.05
10	applications_used.calendar_count	0.03
11	ping_service.wifi_home_percentage	0.03
12	applications_used.weshare_count	0.00
13	applications_used.mobilepay_count	0.00



Conclusive Remarks

- The scope of the study is very limited, inhibiting the ability to make generalizations
- We both confirm and refute links found in previous studies
- We find links not seen in previous studies
- Work hints that it might be possible to computationally model loneliness



Future Work

- Increase generalization of results by including more subjects
- Operationalize resulting classification model to be used as part of real-life applications

Stay Active

