Predicting Spatial Abilities using psychological features

- A Statistical Learning Project -

Grimaldi Francesco

University of Padua,
Master Degree in Data Science

The project

The first aim of the project is to predict the spatial abilities of a person, measured with two psychological tests, using as predictors the gender, some personalities features and some self-assessment wayfinding inclinations.

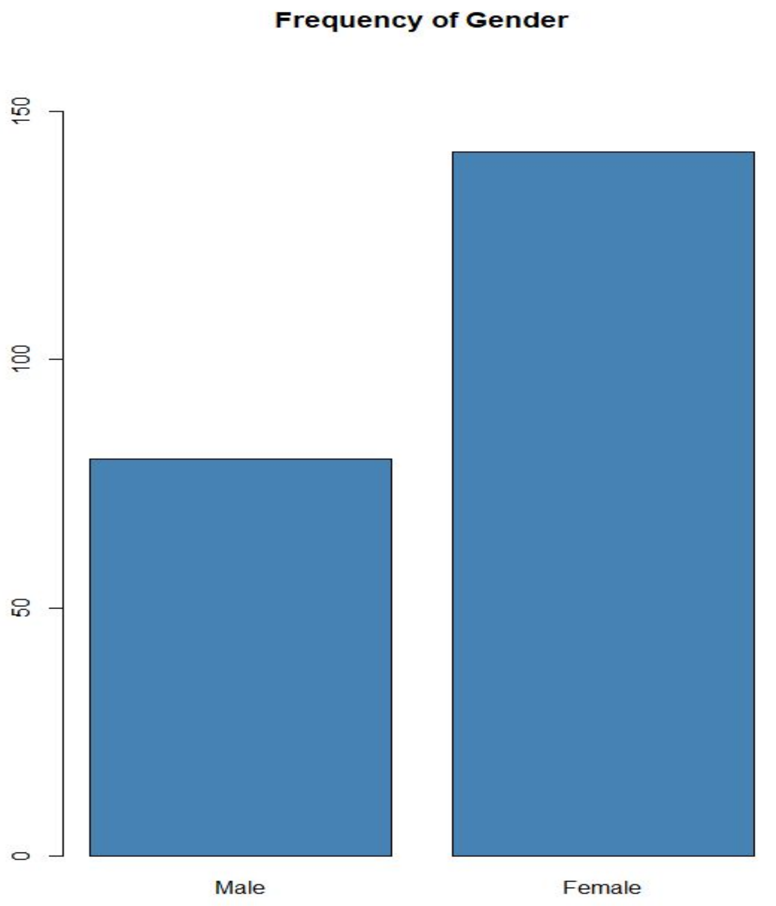
The second aim is to detect which features, from the above ones, are the most useful to predict the spatial abilities.

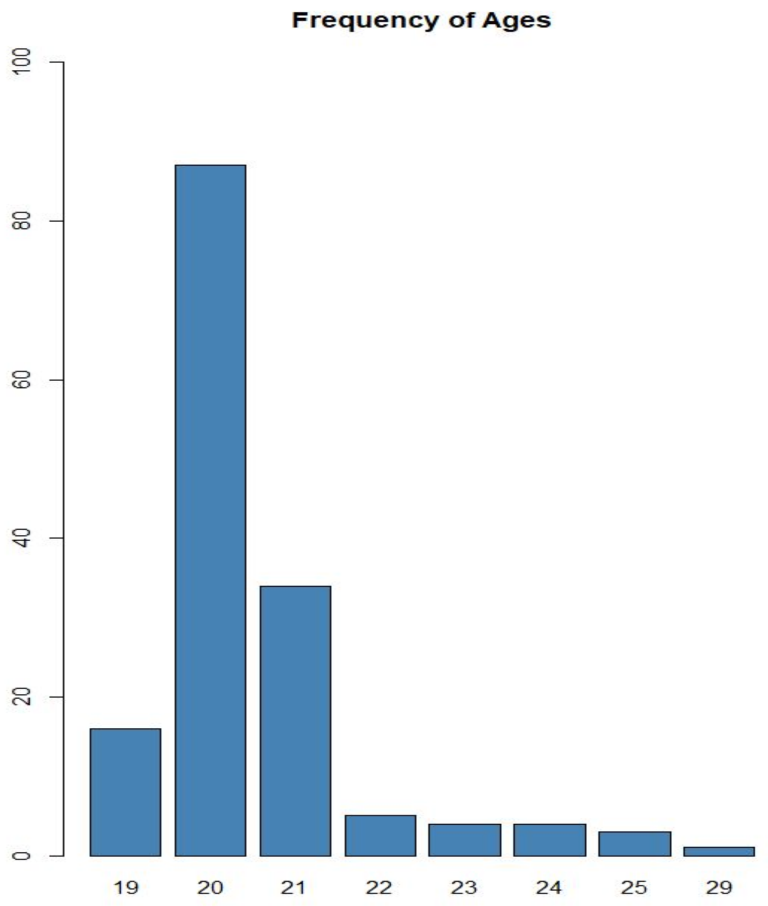
1.
Data Description

The dataset

The dataset contains information of 222 person with the following details:

- All studying at the University of Padua under the Department of General Psychology
- 80 are Male and the remaining 142 are female.
- Around the 95% of the subjects were between 19 and 23 years old.


Frequency of Gender
150
100
50
0
Male
Female


Frequency of Ages
100
80
60
40
20
0
19
20
21
22
23
24
25
29

Description of predictors: Big Five Questionnaire

Informations regarding personality were taken using the Big Five Questionnaire (B5Q, version Caprara et al, 2002), probably the most famous personality test used today.

It consist of 134 items about traits of personality, these items are on a Likert Scale (discrete scale) ranging from 1 to 5. The scale of the measure is ordinal.

The items are divided in 10 categories, called facets. Each two facets form a trait.

B5Q: The list of Facets and their Traits

1. **Dynamism and Dominance → Extraversion**
2. **Cordiality and Cooperativeness → Agreeableness**
3. **Scrupulosity and Perseverance → Conscientiousness**
4. **Emotion Control and Pulse Control → Emotional Stability**
5. **Openness to Experience and Openness to Culture → Openness**

B5Q: Example of some items

	Quite Often	Often	Sometimes	Rarely	Almost Never
	1	2	3	4	5
1. I feel like I'm on an emotional roller coaster.	○	○	○	○	○
2. During tough times, I am more prone to unhealthy behaviors (abusing drugs or alcohol, eating unhealthy foods, getting less sleep).	○	○	○	○	○
3. I feel uneasy in situations where I am expected to display physical affection.	○	○	○	○	○
4. I present myself in ways that are very different from who I really am.	○	○	○	○	○
5. I procrastinate on matters relevant to work.	○	○	○	○	○
6. I break promises.	○	○	○	○	○
7. I lose important things/documents.	○	○	○	○	○

Description of predictors: Navigation Aid Questionnaire

Three features coming from three items on a Likert Scale from 1 to 6, the NAD (adapted by Munzer, Zimmer, Schwalm, Baus, Aslan, 2006) investigates what a person use during navigation tasks.

The first item ask if a person uses a map, the second item ask if a person uses a GPS and the third one ask if a person tends to ask for verbal indication.

Description of the predictors: Self Assessment wayfinding inclinations

For this scope a series of self assessment questionnaires has been used and a total of six features have been extracted:

1. Space Anxiety Scale: Assess anxiety during behaviours involving space reasoning.
2. QACO-knows: Assess the preference of a person to stay in familiar places
3. QACO-exploration: Assess the pleasure in exploring new places
4. Cardinal Points: Assess how much a person tends to use Cardinal Points during space-orientation tasks
5. Visual and Route Style: Assess the preference to use naive orientation style
6. Sense of Direction: Assess your confidence to not to get lost.

Factors of Self-Assessment wayfinding inclinations

In the dataset are present two features called “Positive Factor” and “Negative Factor” which comes from a previous factorial analysis on the data.

The “Positive Factor” is derived by the union of the features: QACO-exploration, Cardinal Points, Visual and Route Style and Sense of Direction.

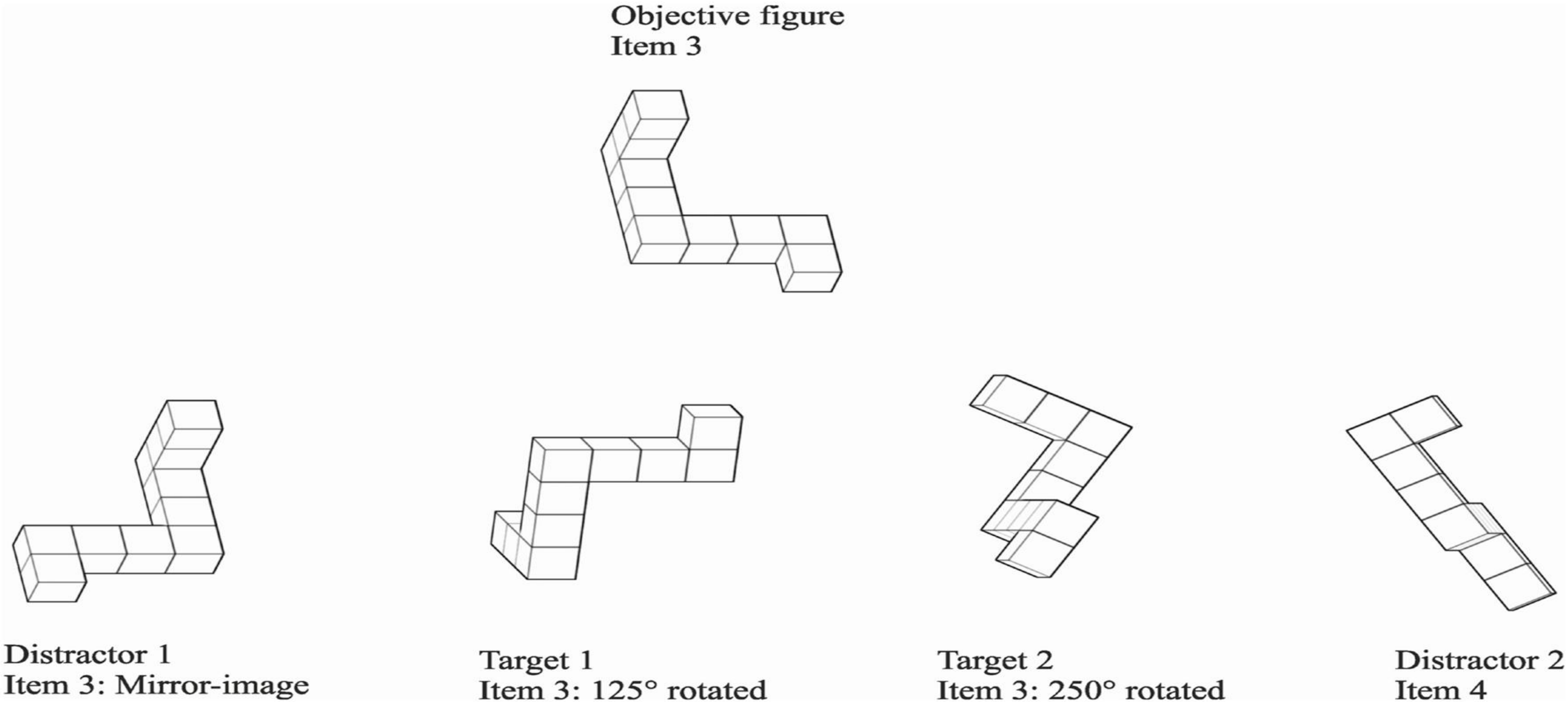
The “Negative Factor” is derived by the union of: QAS and QACO-known.

Response Variables: Mental Rotation Task (MRT)

The Mental Rotations Test (MRT; adapted from Vandenberg, Kuse, 1978; De Beni et al., 2014). It consists of 10 items each one requiring in identifying two of four abstract 3D objects matching a target object but in a rotated position.

The maximum time given is 5 minutes. For each item a point is given if both alternatives were detected (Maximums score: 10)

Example of MRT item



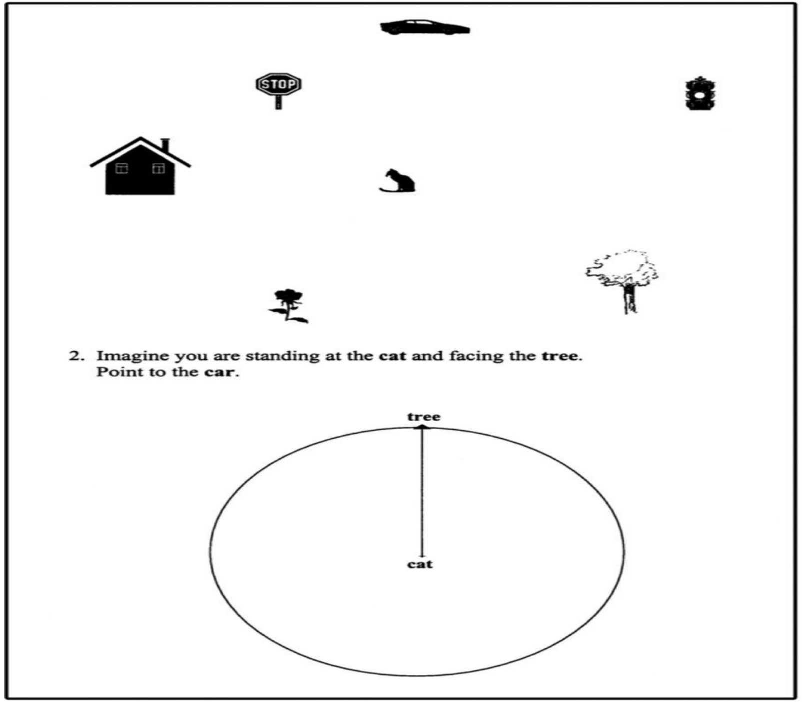
Point Perspective Taking Task (PTT)

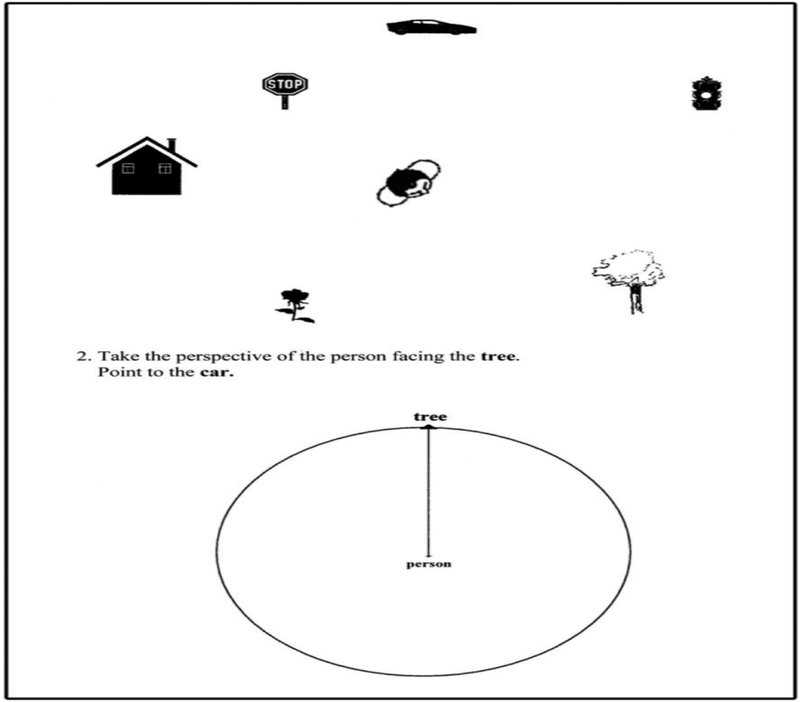
The Point Perspective Taking task (PTT; adapted from Kozhevnikov and Hegarty, 2001; De Beni et al., 2014). It consists of 6 items.

The maximum time given is 5 minutes.

It is calculated the degree of difference between the angle individuated and the correct one and the sum of the degree of differences is made (i.e. higher score high number of errors).

Example of two PTT items





2.
Clean and Filter Data

Removing misleading information

After importing the dataset, it was noticed that two extra rows, representing the mean and the standard deviation of every features and several extra columns, filled with NA, were present.

So, the first action was to remove these extra columns and rows.

Checking variables format

Next step was to check how our variables were treated from the R environment.

Every variables was treated as numerical and this was OK for us apart from the variable Gender which is categorical. For this reason a transformation to factor was applied, where the number “1” was given to the Males and the number “2” was given to the Females.

Dealing with missing values (NA)

A search for any missing values was performed revealing that seven NA were present in the dataset. All these seven values were about the PTT scores of seven people.

Since the low number of instances, it was decided to deal with these NA values by substituting them with the mean of the PTT.

Collinearity: removing redundant information

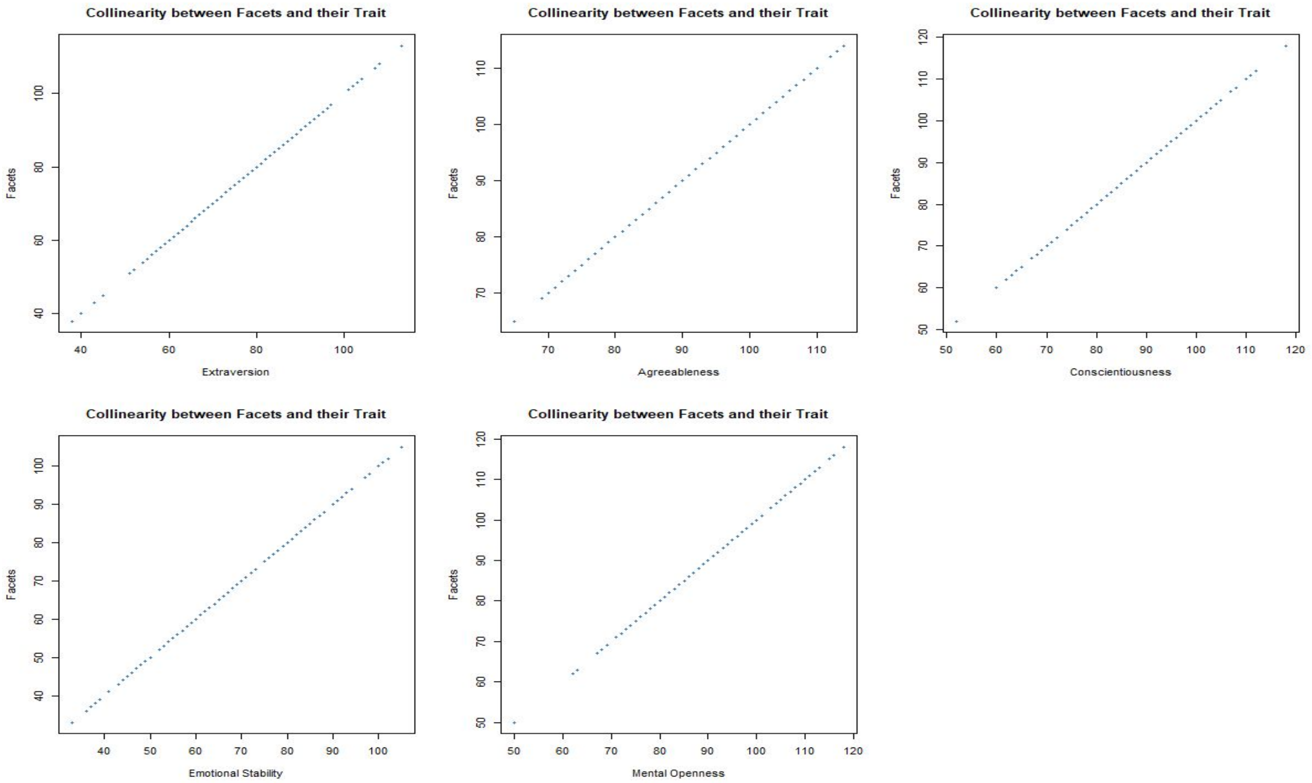
Collinearity between some features can be interpreted as the fact that this two features perfectly correlate.

In our case this effect is present in two cases:

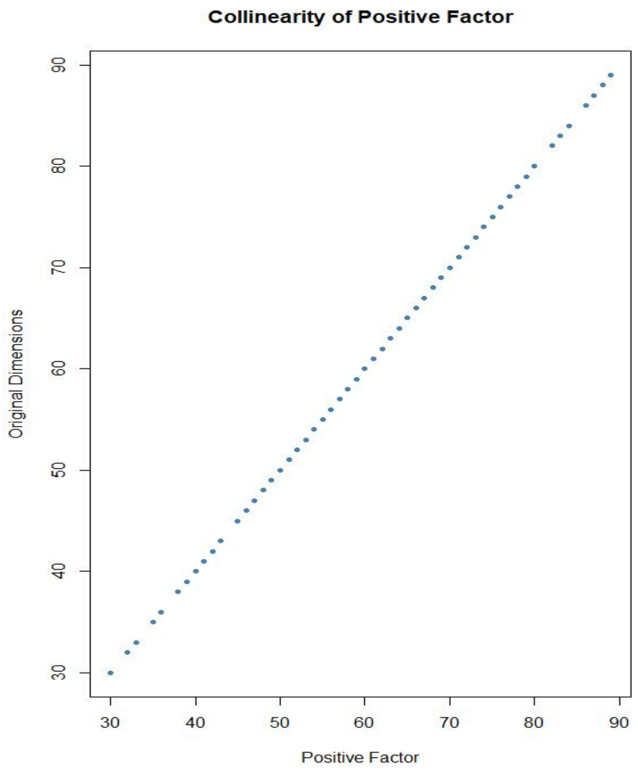
1. Between the sum of the facets and their corresponding trait
2. Between the “Positive Factor” and “Negative Factor” and the sum of their corresponding dimensions.

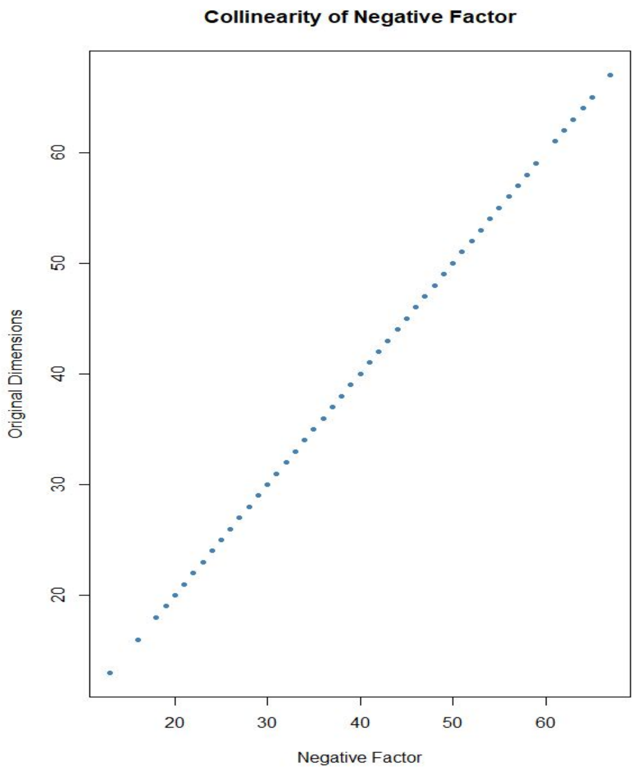
For this reason it was decided to remove the five B5Q’s traits and the two factors.

Collinearity between Facets and Traits



Collinearity between “Factors” and their dimensions



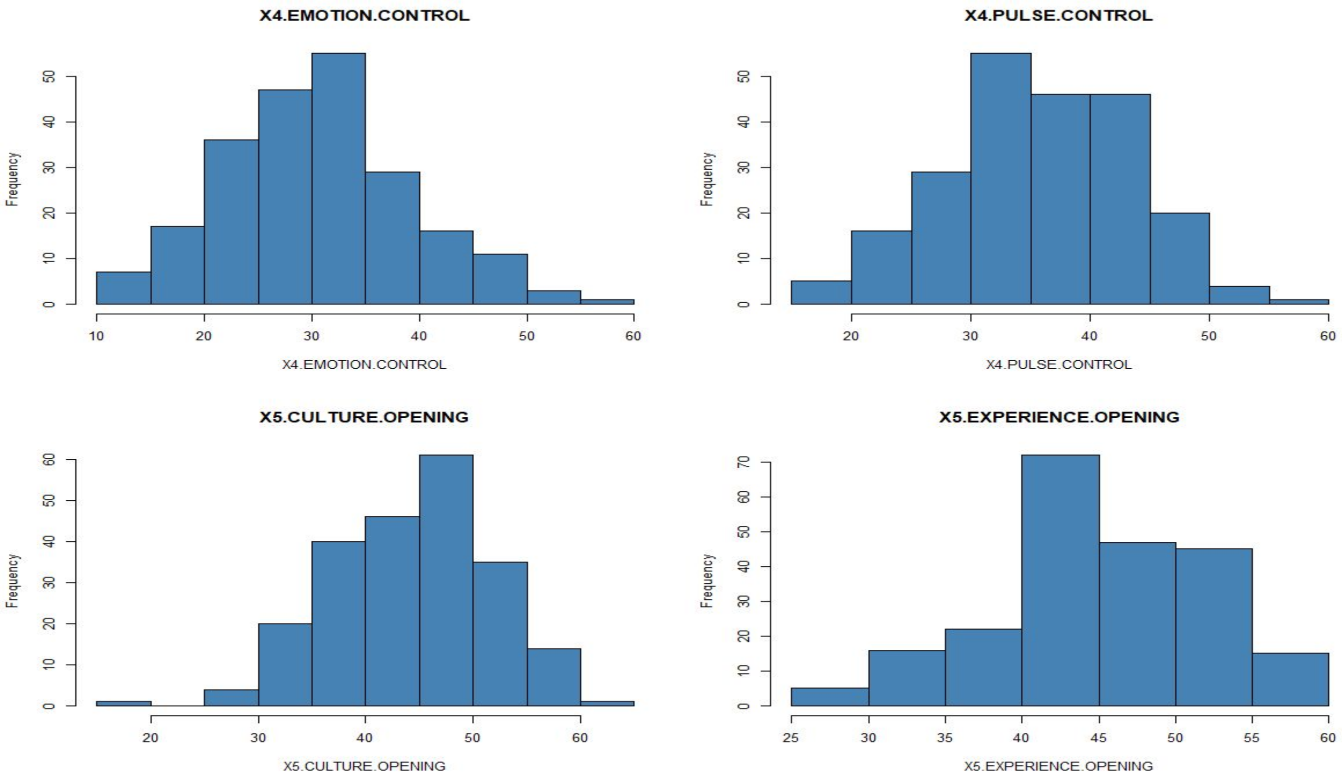


3.
Exploring the Data

Checking Distribution

After a brief look at descriptive statistics index of the features (e.g. Mean, Standard Deviation, Quartiles etc.) a graphical representation of the distribution of the features was done by using histograms for the numerical values and barplot for the case of categorical features (Gender in our case).

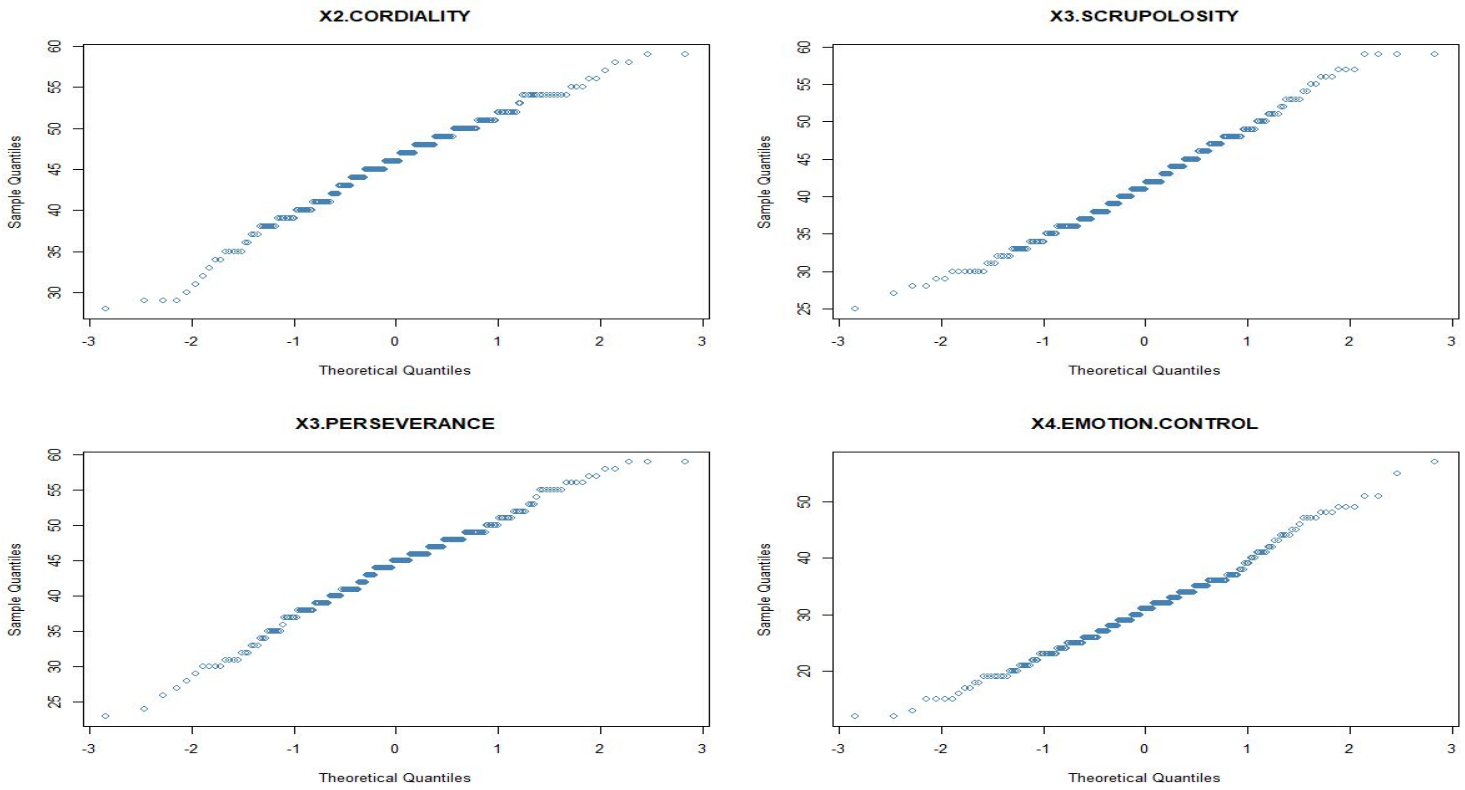
Distribution of some features



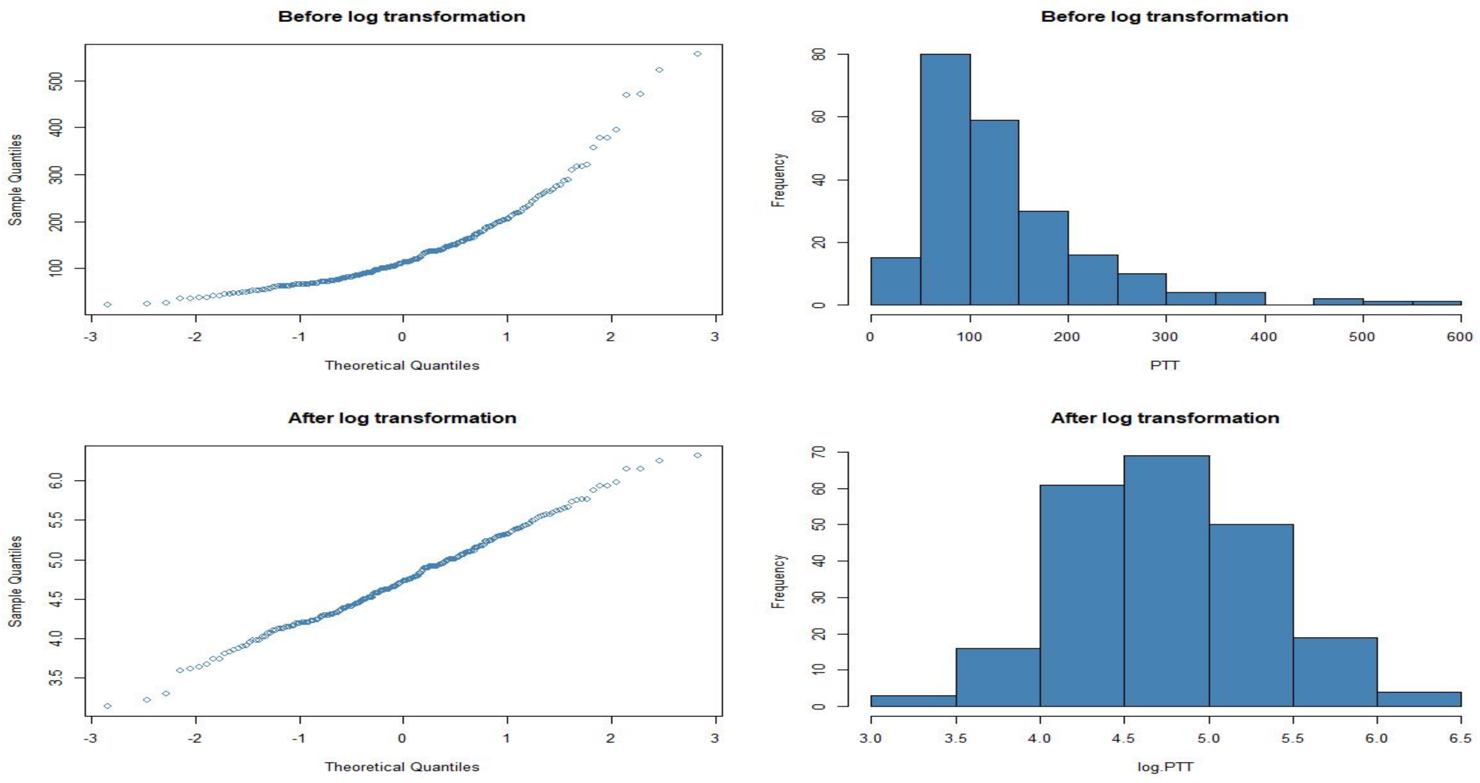
Checking normality of the data

Since a lot of statistical methods and model have a priori assumption that the given data comes from a normal distribution, a check on the normality of the data has been done by using graphical representation.

Normality of the features



A non normal distribution: PTT to logarithm of PTT

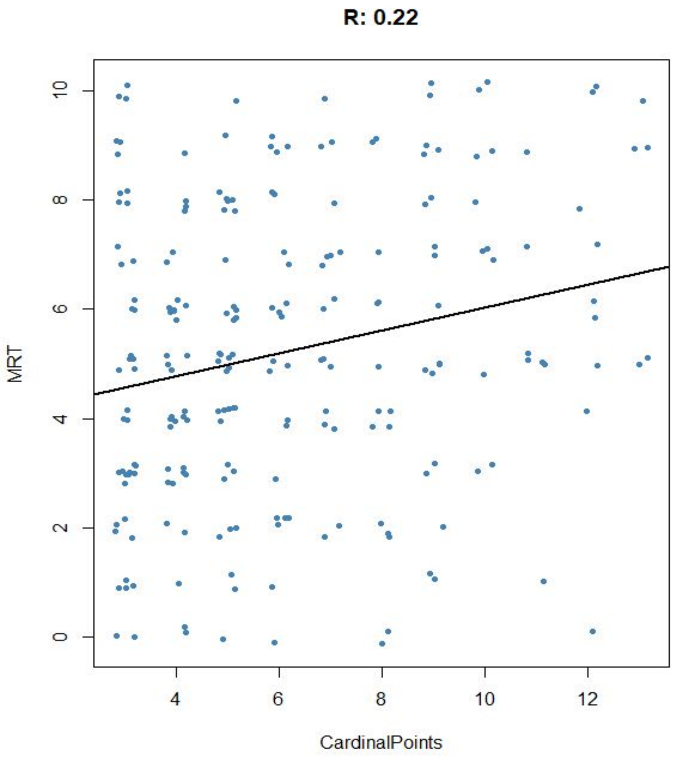


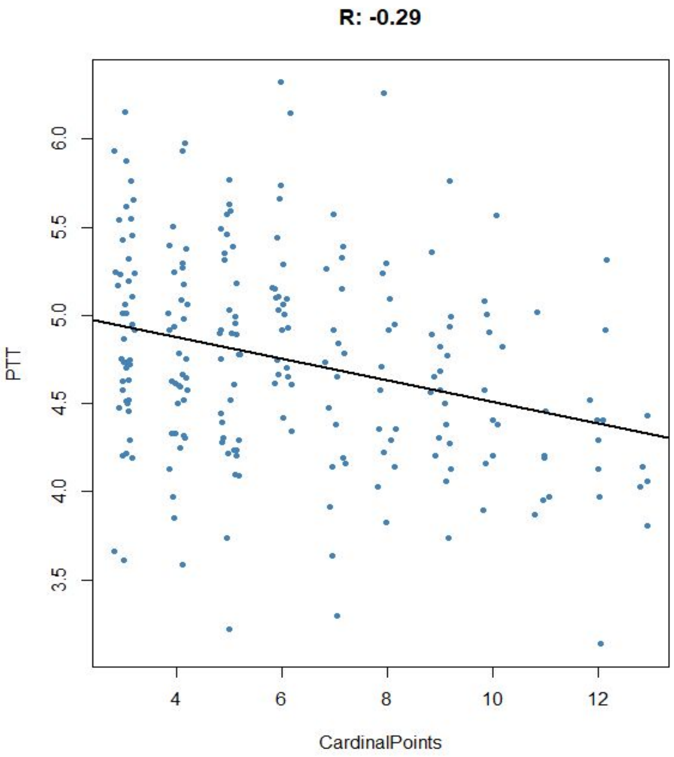
Relation of MRT and PTT with the other features

In order to see the relations of MRT and PPT with the predictors, some graphical representation and correlation were performed.

For every features and our response variables it was plotted the data distributions, with line representing the linear regression and the information of the correlation between the two. For the case of Gender it was used a boxplot

The strongest relations: Cardinal Points





PTT and MRT: from numerical to binary variables (1)

There are few reasons why it was decided to transform our response variables from numerical to binary:

1. As we have seen the strongest correlation was r = -0.29 and all the other features had so very weak correlation that the mean of the correlations for MRT and PTT was 0.02 and -0.04 respectively. These observation made us think that a regression model would have been really bad, meanwhile a classification problem could have given us better results.
2. Other than this practical reason we had also a theoretical one. In fact, the aim of the project was to assess the spatial abilities of a person (e.g low and high spatial ability) and not to assess a score of task which per se isn't meaningful

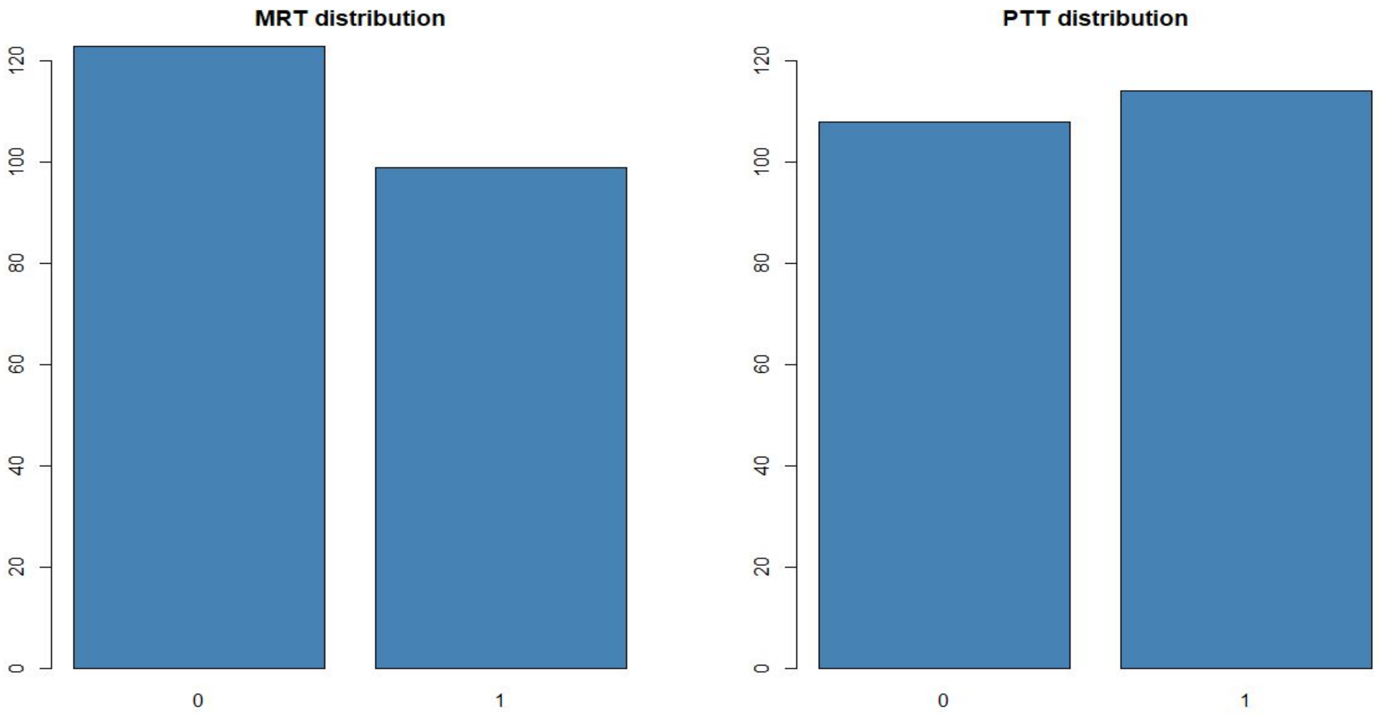
PTT and MRT: from numerical to binary variables (2)

Since both MRT and PTT (after log transformation) showed a normal behaviour it was decided to do the followings two procedures:

1. for MRT to assign “1” to the values higher than the mean and “0” to the values below or equal. So that “1” meant score showing good result and “0” score with bad result.
2. for PTT, since high scores meant bad results, to assign “0” to the values higher than the mean and “1” to the values below or equal. So that “1” meant score showing good result and “0” score with bad result.

This gave us almost balanced group (MRT: 123 vs 99; PTT: 108 vs 114).

MRT and PTT distributions

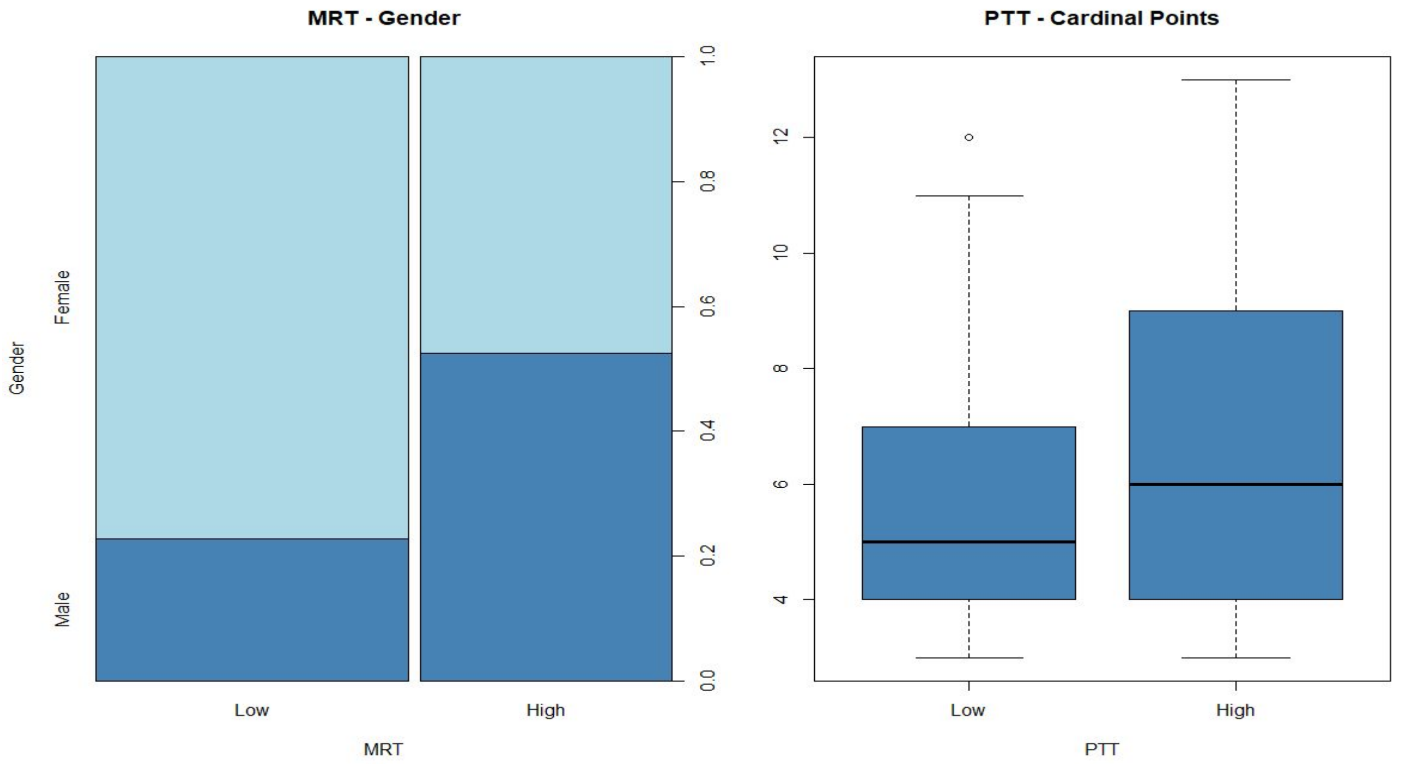


Relations of binary MRT and PTT with the predictors

To explore the relations of our transformed response variables others graphical representation were used:

Since MRT and PTT now are categorical variables boxplot were used to explore the relations, meanwhile for the case of gender it was used a Stacked Columns Charts.

Gender for MRT and Cardinal Points for PTT



4.
Models

Classification

What we want is to try to predict if the PTT and MRT will be 0 or 1 based by the information of our twenty predictors.

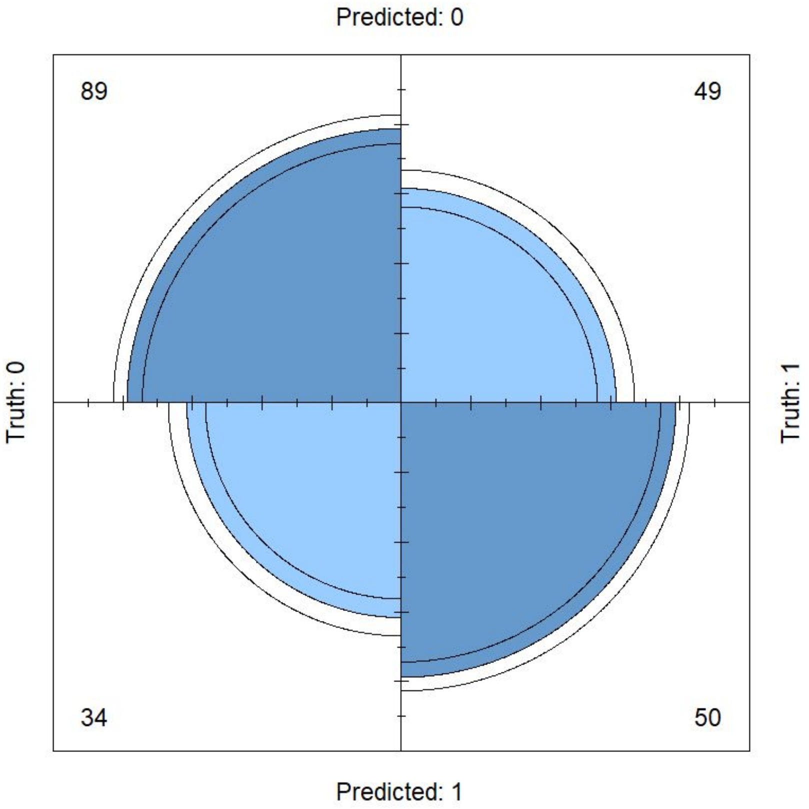
For this reason we will used supervised learning classification algorithms that, given a set of labeled data, will try to learn about pattern and distribution of our response variables.

Measures used to evaluate a model

To evaluate the goodness of a model a series of metrics were used:

1. Accuracy (tp/n + tn/n): It computes the proportion of times the classification guessed right the instances.
2. Sensibility (tp/P): It measures the proportion of times the model guessed right the instances labeled as “1”.
3. Specificity (tn/N): the same of Sensibility in the case of “0” examples.
4. Receiving Operating Characteristic Curve (ROC curve): its a curve given by the ratio between Sensibility and 1-Specificity at various threshold.
5. Area Under the Curve (AUC): computed as the area under the ROC Curve.

Confusion Matrix



A confusion matrix is table useful to visualize the result of a classification model:

- In the top-left corner we find the number true negative examples (tn)
- In the top-right corner we find the number of false negatives (fn)
- In the bottom-left corner we find the number of false positives (fp)
- In the bottom-right we find the number of true positives (tp)

Validation Procedure

A Leave One Out Cross Validation (LOO-CV) has been used in all the model.

LOO-CV consist in using only one instances as test set and the others as training.

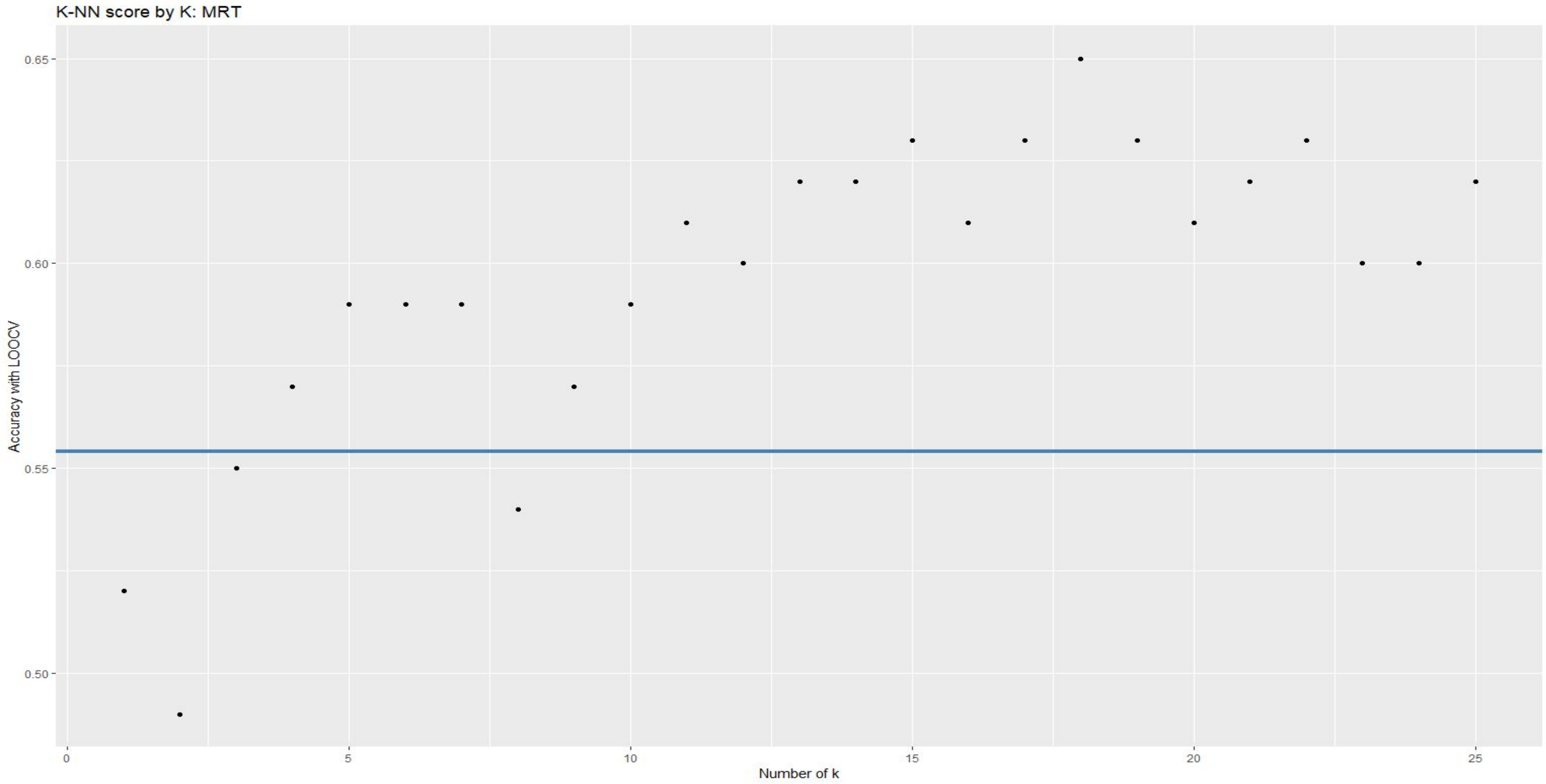
This procedure is done for every instances and then the metrics are the means of the metrics of all the models

Model #1: K-NN

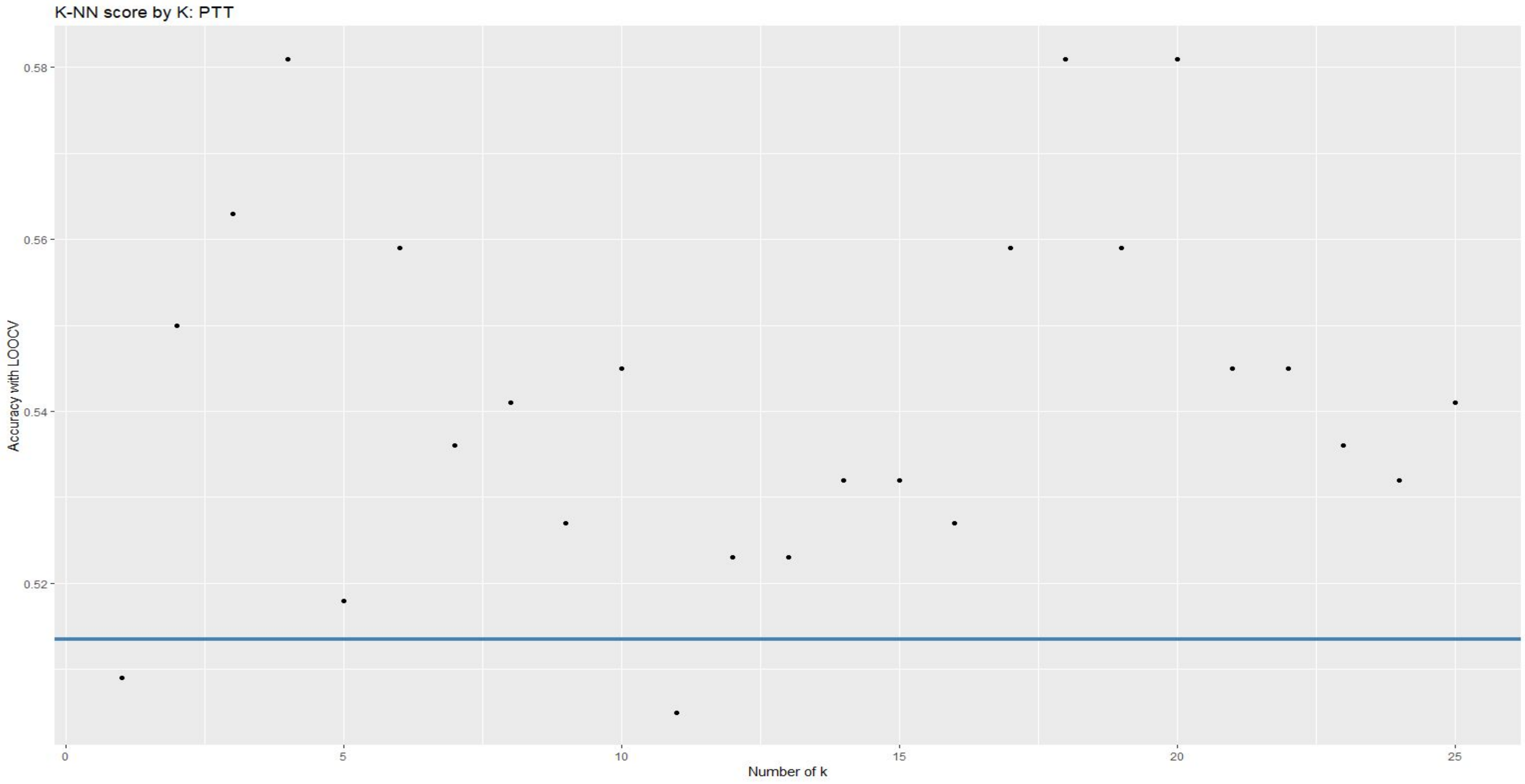
This first model is non-parametric and classify a instances by computing what is the most present class in the neighborhood (the k-closest instances) of that example (see report to further details).

The accuracies were low for both MRT and PTT (0.653 and 0.581 respectively). See report for all the metrics

Results of K-NN for MRT by various K



Results of K-NN for PTT by various K

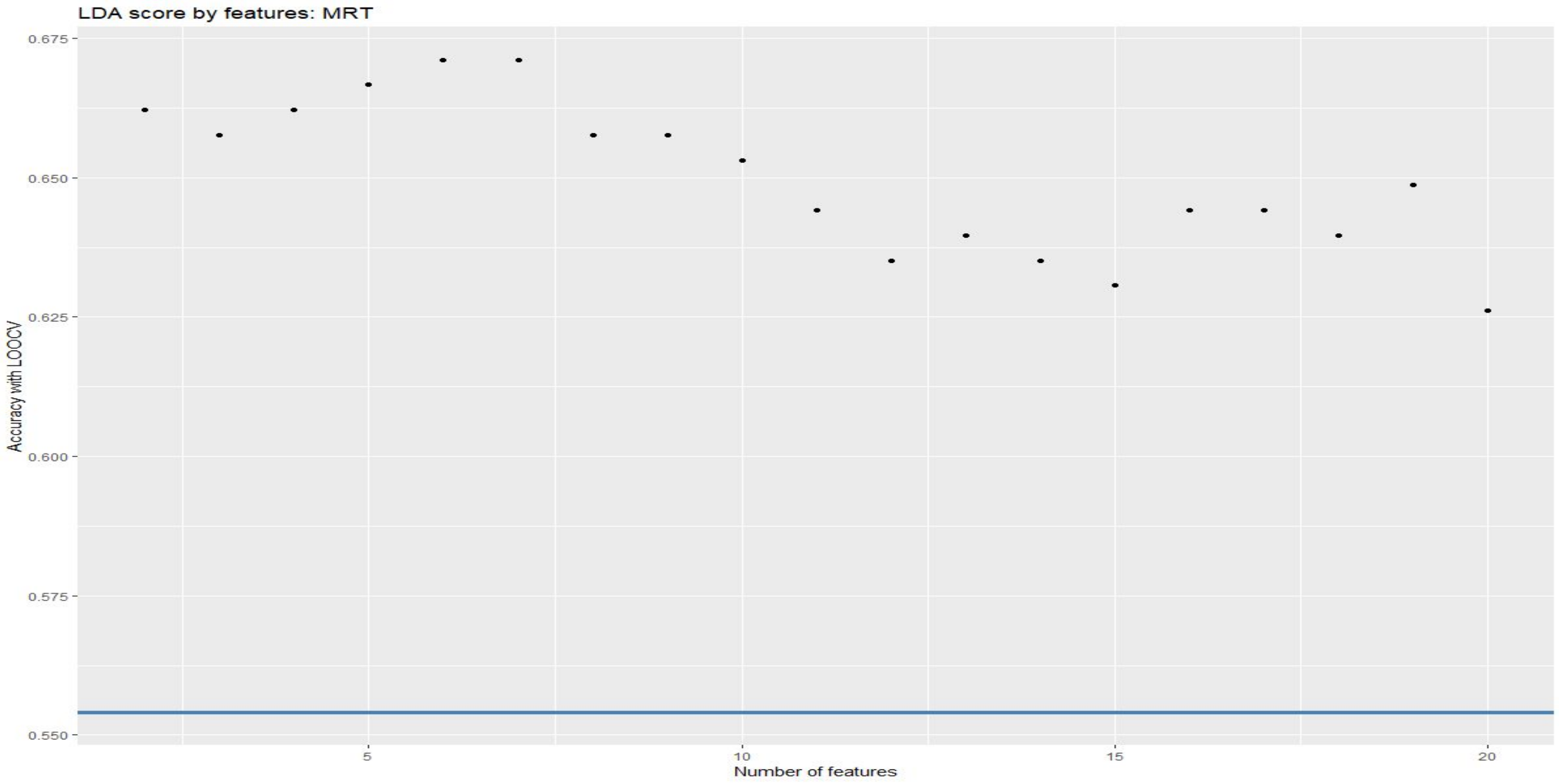


Model #2: Linear Discriminant Analysis (LDA)

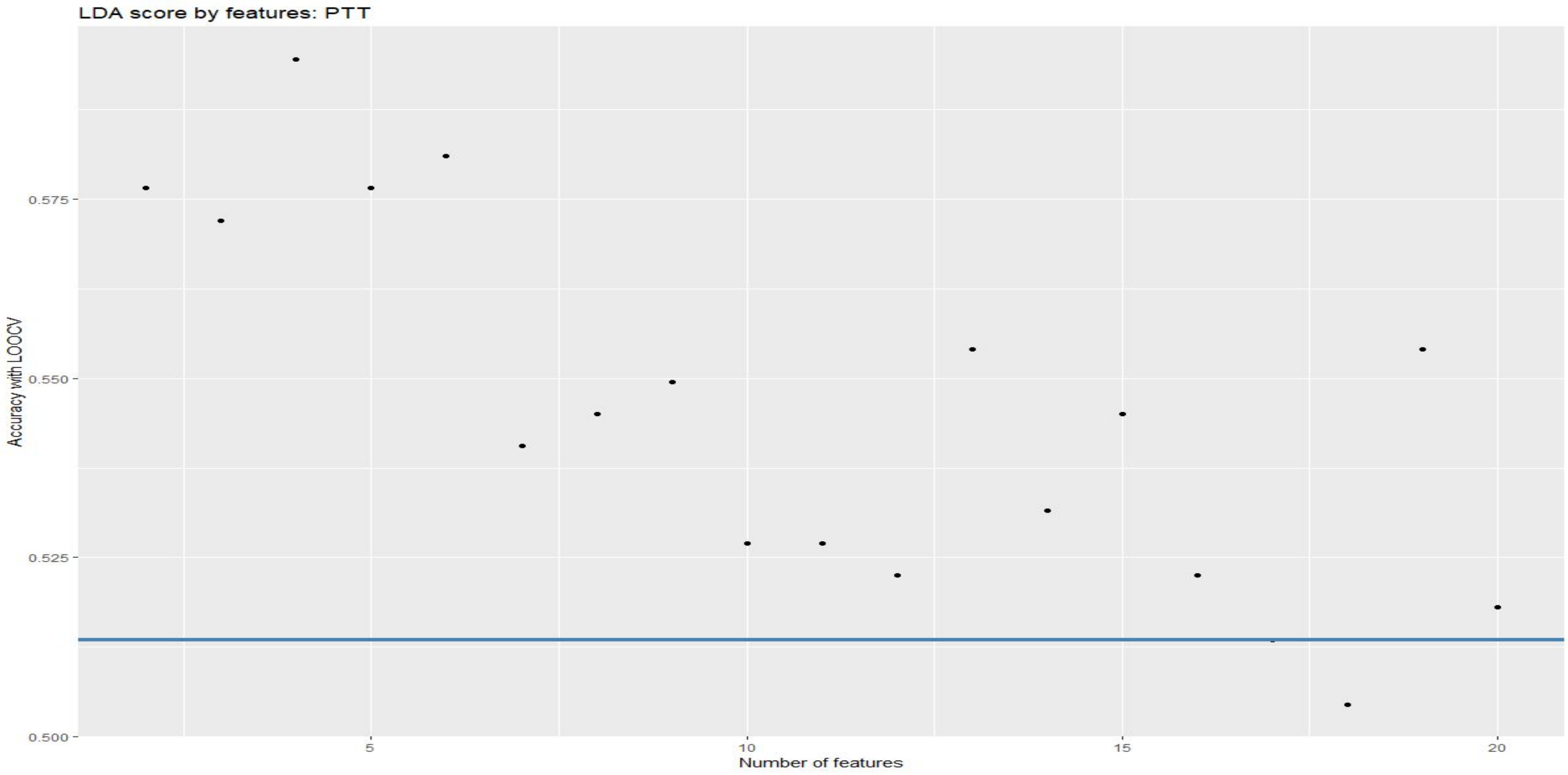
LDA is an algorithm for classification based on estimating the means of the distributions of the classes and then classify an example based on the probability of being generated by that random variable having that distribution. It’s also take in consideration the prior distributions (See report to see how the algorithm works).

Since LDA models with all the features for both MRT and PTT weren’t good a step-forward feature search based on the relative distances of the means of the two class were performed. (See report to see how the procedure works)

Accuracy of MRT by changing the number of features



Accuracy of PTT by changing the number of features

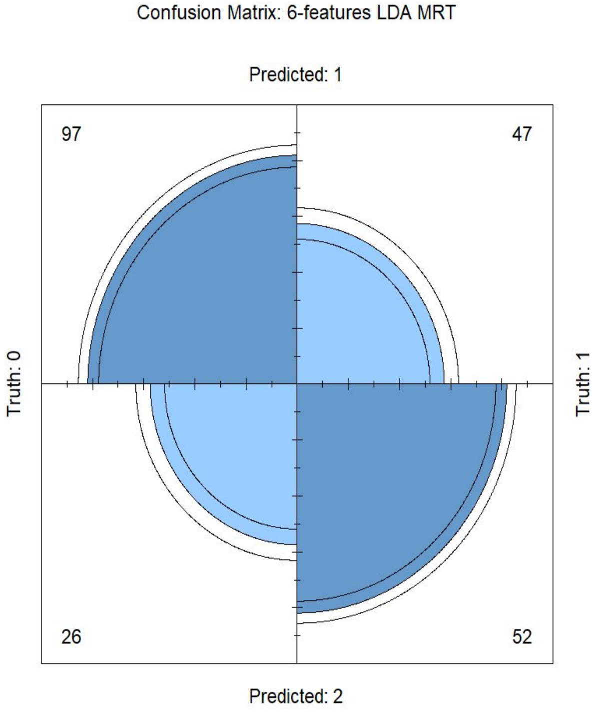


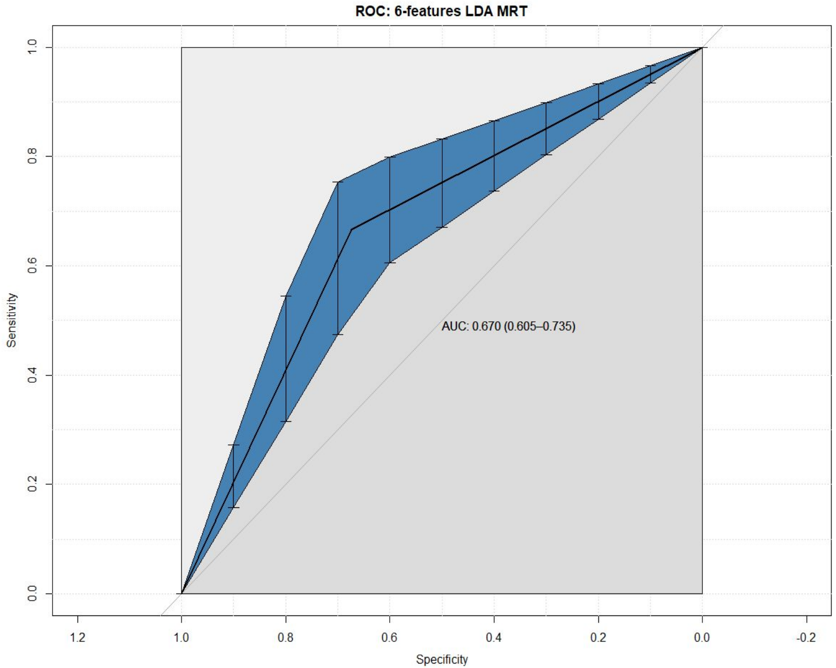
Result and best features of LDA for MRT

1. Gender
2. Emotion Control
3. Pulse Control
4. QAS
5. Cardinal Points
6. Sense of Direction

1. Accuracy: 0.671
2. Sensitivity: 0.526
3. Specificity: 0.789
4. AUC: 0.670

Result of LDA for MRT with six features



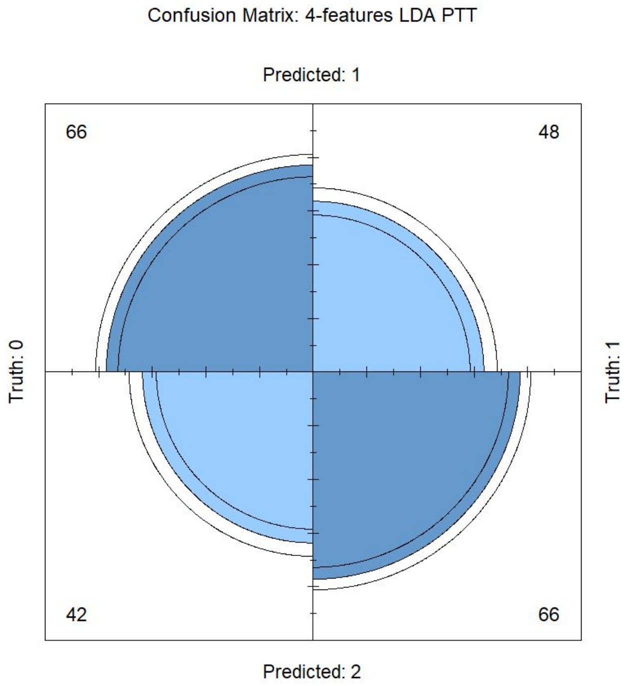


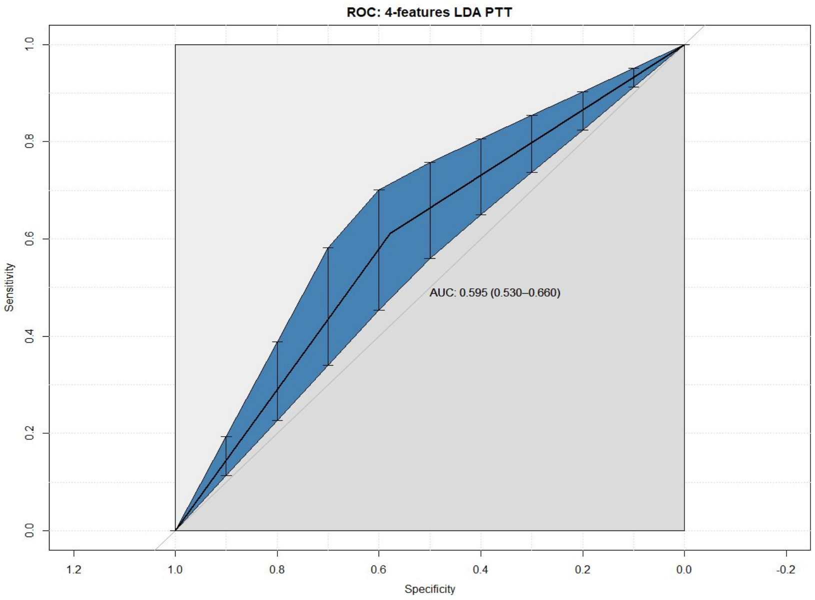
Result and best features of LDA for PTT

1. Gender
2. Map Use
3. Verbal Indication
4. Cardinal Points

1. Accuracy: 0.595
2. Sensitivity: 0.579
3. Specificity: 0.611
4. AUC: 0.595

Result of LDA for PTT with four features:





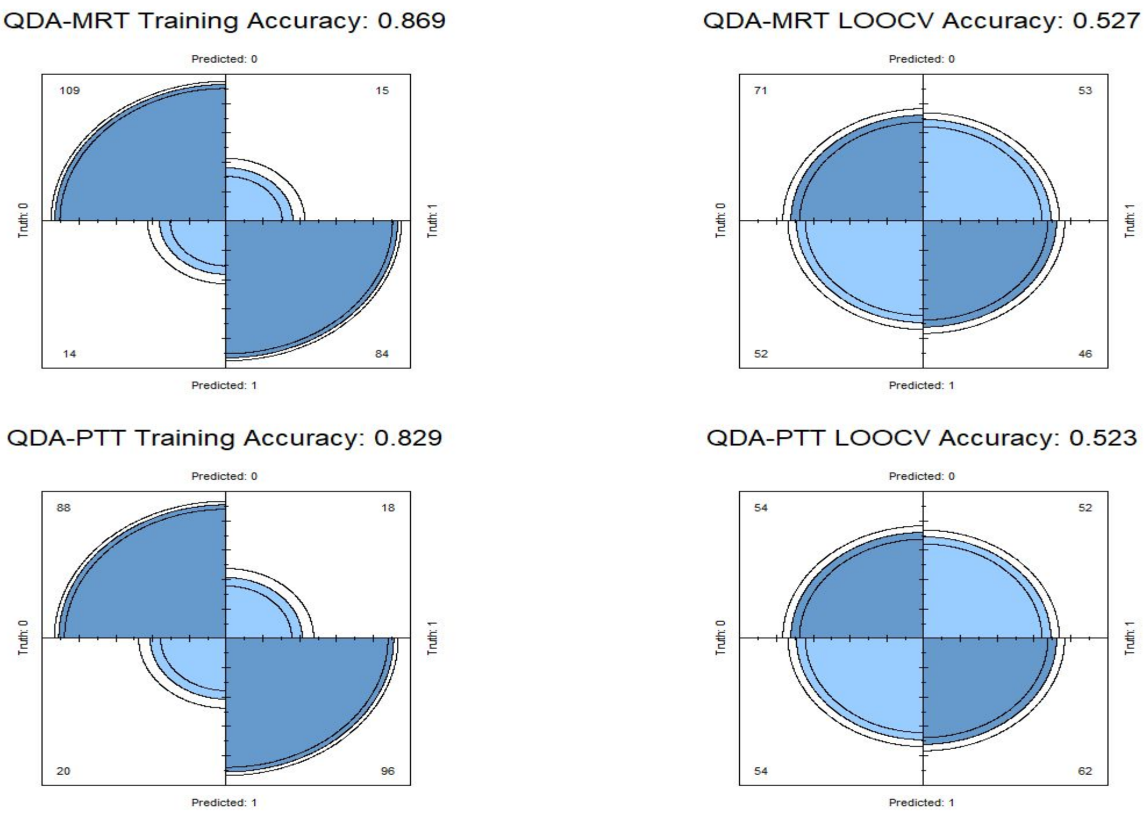
Model #3: Quadratic Discriminant Analysis (QDA)

QDA is a generalization of LDA, where other then estimating the means we estimate also the covariance matrices, one for each class. It can give non linear boundaries: this could be good when our data distribution isn’t linear but can also result in overfitting and catching random noise.

The procedure of features search was the same of the case of LDA but it brought worse result for both PTT and MRT.

To see further details of the algorithms and of the results look at the report.

QDA overfitting: train vs validation accuracy



Model #4: Generalized Linear Model (GLM)

As the name suggest GLM is a generalization of Linear Model, when this one is not the best for some kind of problems.

In Linear Regression we say that the mean of our response variable (E[Y]) is computed as βX, where X is the values of our features and β is the values of the coefficients.

In Binary problems where E[Y] is the probability to belong to the positive class, the standard linear regression function doesn’t suite our problem since it can take values outside the interval [0,1]. For this reason we use a link function that change our equation.

Linear Regression vs Logistic Regression

In linear Regression as we previously said we have that:

$$E[Y] = \beta X$$

When we use a GLM with link function:

$$\log(x/(1-x))$$

We say that we are using a Logistic Regressor that give us:

$$\log(E[Y]/(1-E[Y]) = \beta X \Longrightarrow E[Y] = \exp\{\beta X\}/(1+\exp\{\beta X\})$$

(See further detail in the report)

Step-Forward Features Selection for GLM

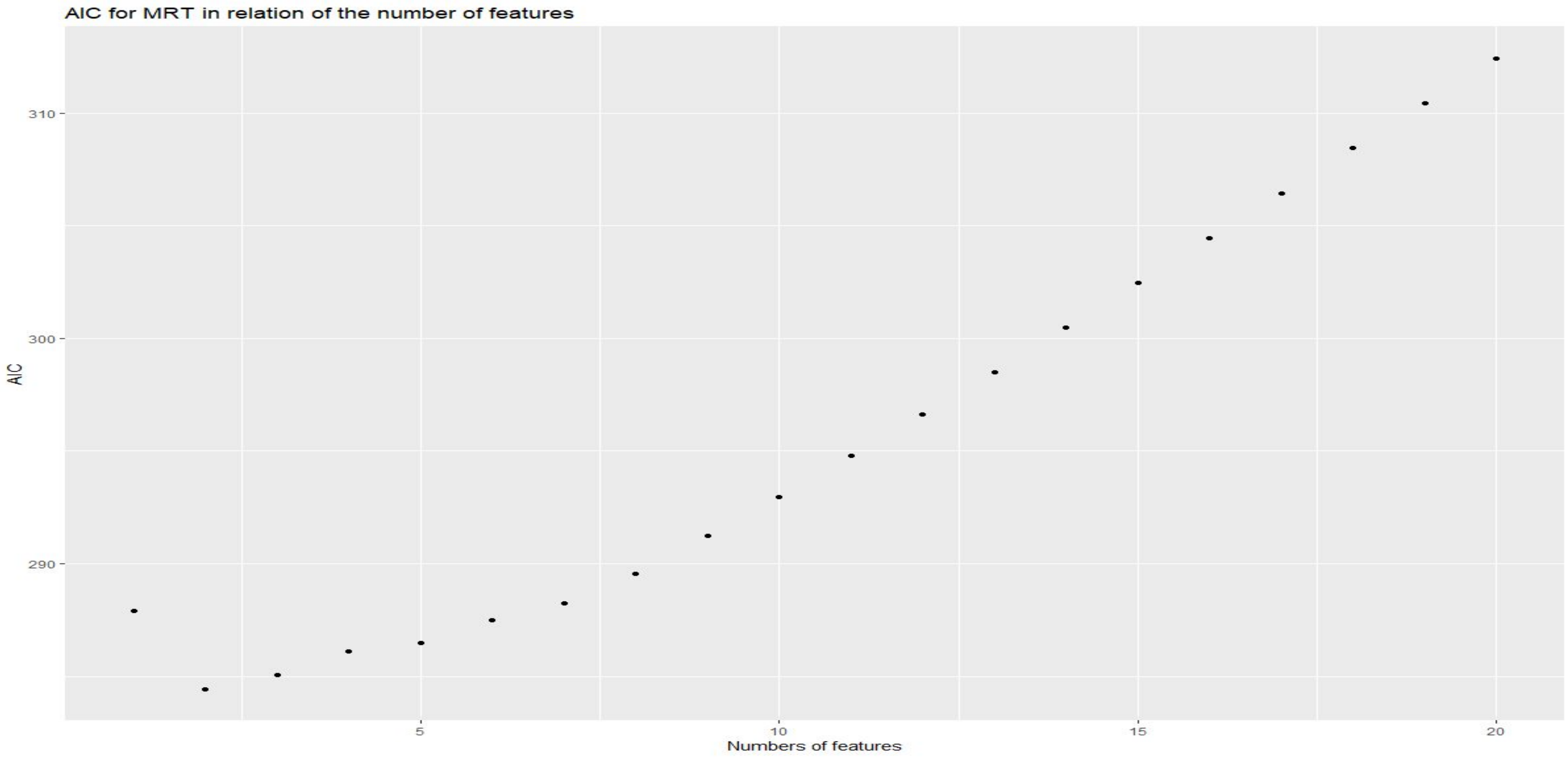
Once again a feature selection has been performed to search for the best model possibles (to see full details see the report).

In order to select the best features and model the Aikannen Information Criterion (AIC) was used. The AIC takes into consideration the goodness of fit (L) and the number of features (k) in this way:

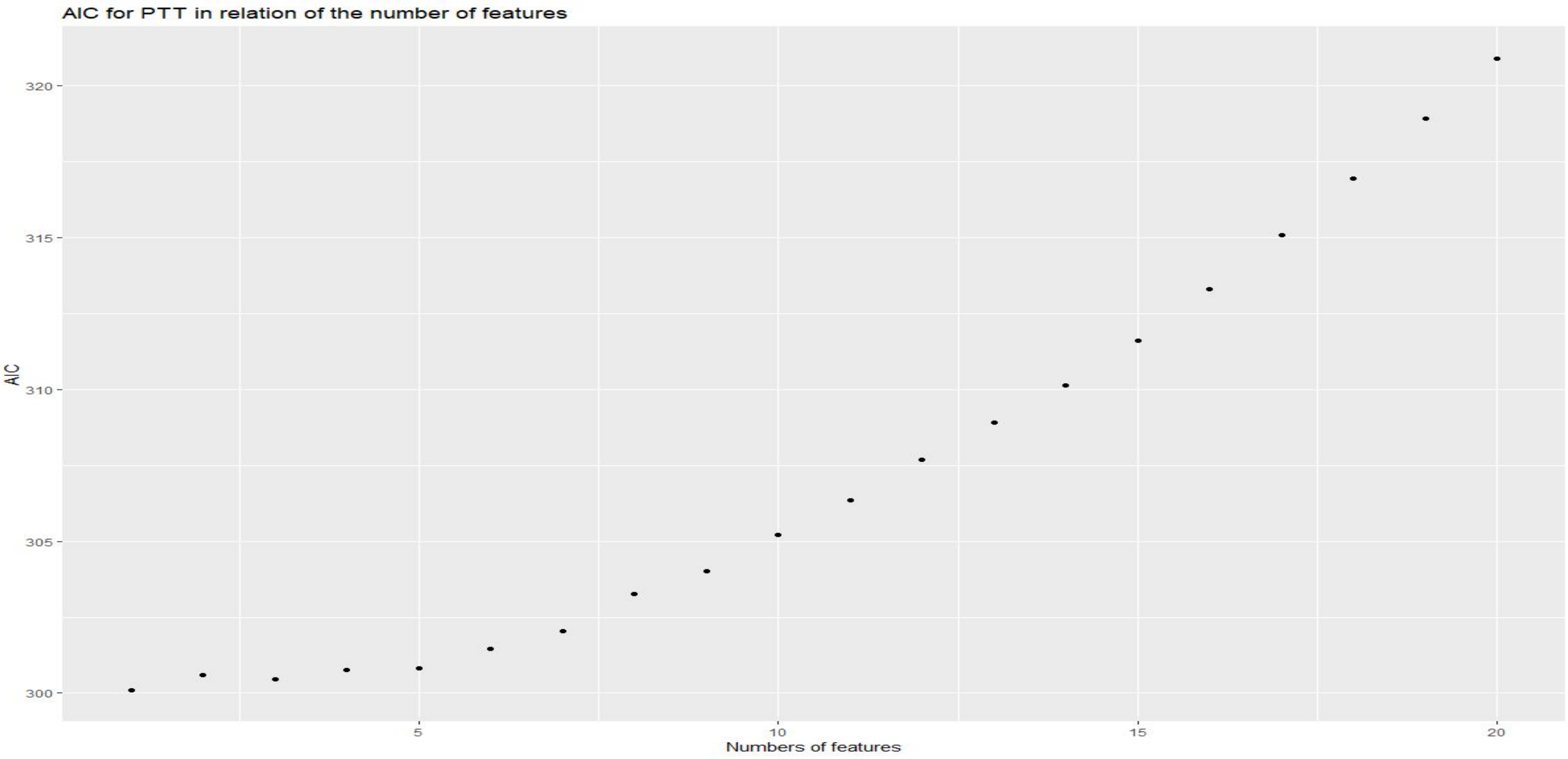
$$AIC = 2*k - 2*\ln(L)$$

Be aware that the AIC is a comparative index so that the model with the lowest AIC is considered the best.

AIC by number of features in the case of MRT



AIC by number of features in the case of PTT

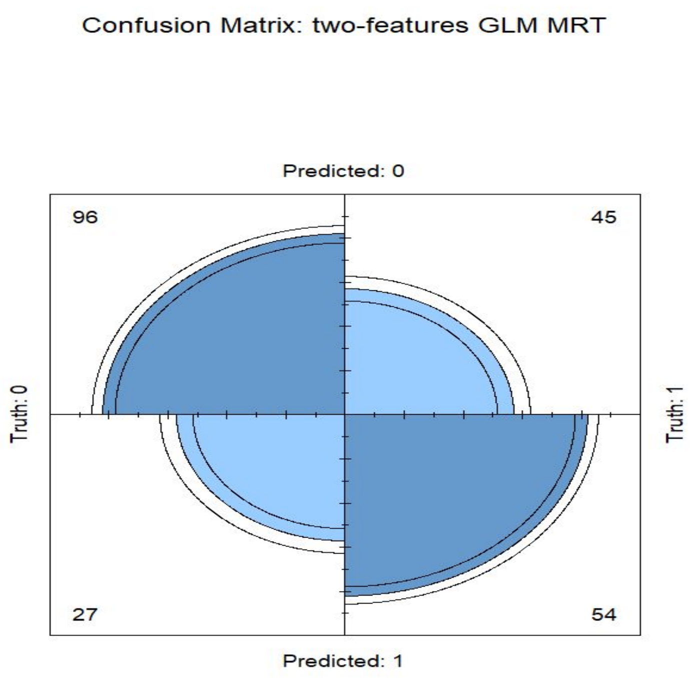


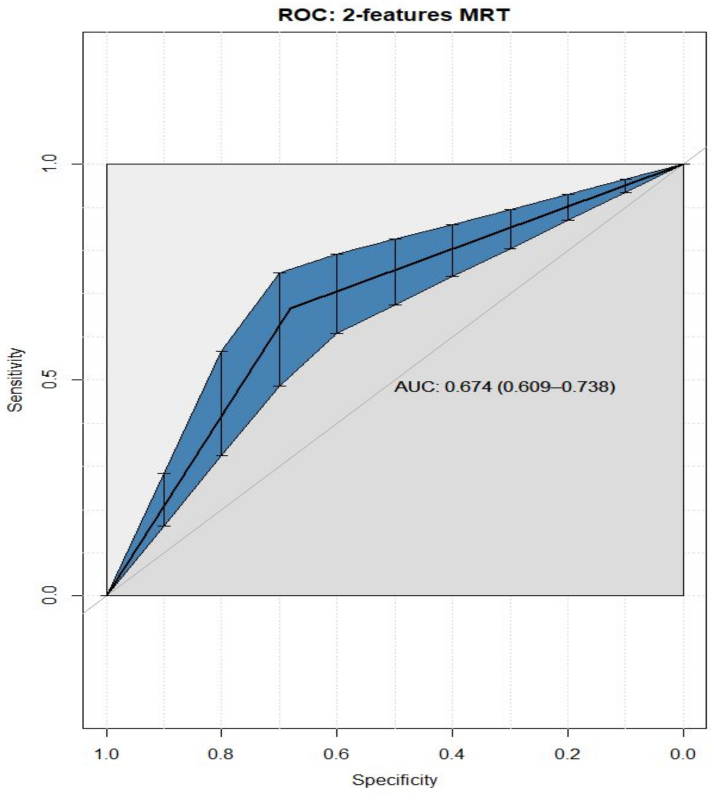
Result and best features of GLM for MRT

1. Gender
2. Emotion Control

1. Accuracy: 0.676
2. Sensitivity: 0.535
3. Specificity: 0.789
4. AUC: 0.674

Confusion Matrix and ROC Curve for GLM-MRT



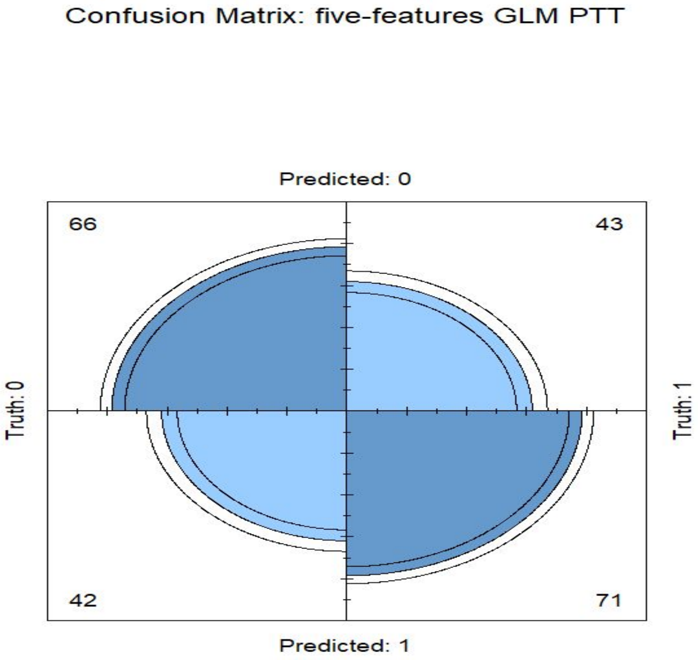


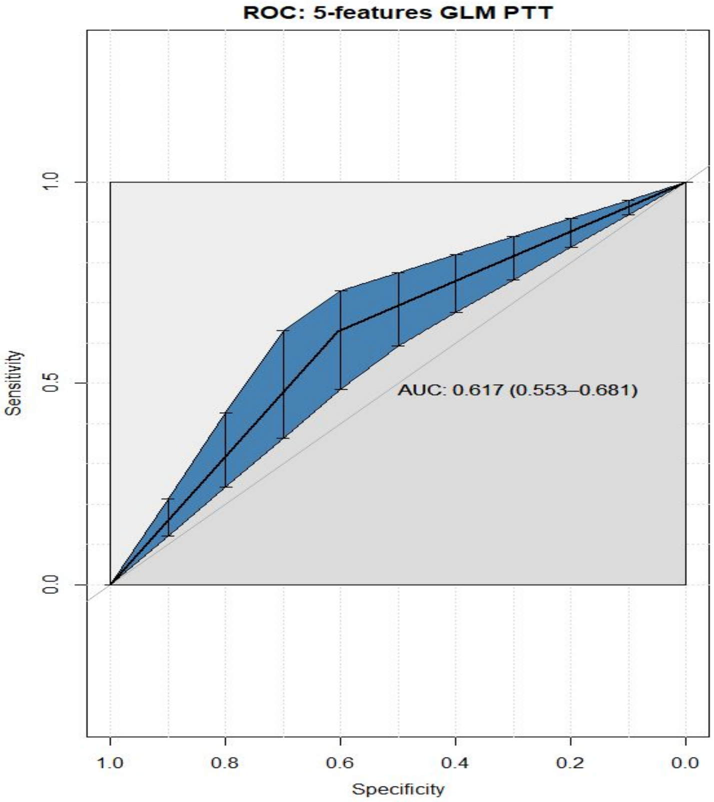
Result and best features of GLM for PTT

1. Cardinal Points
2. Emotion Control
3. Experience Opening
4. Culture Opening

1. Accuracy: 0.617
2. Sensitivity: 0.623
3. Specificity: 0.611
4. AUC: 0.617

Confusion Matrix and ROC Curve for GLM-PTT



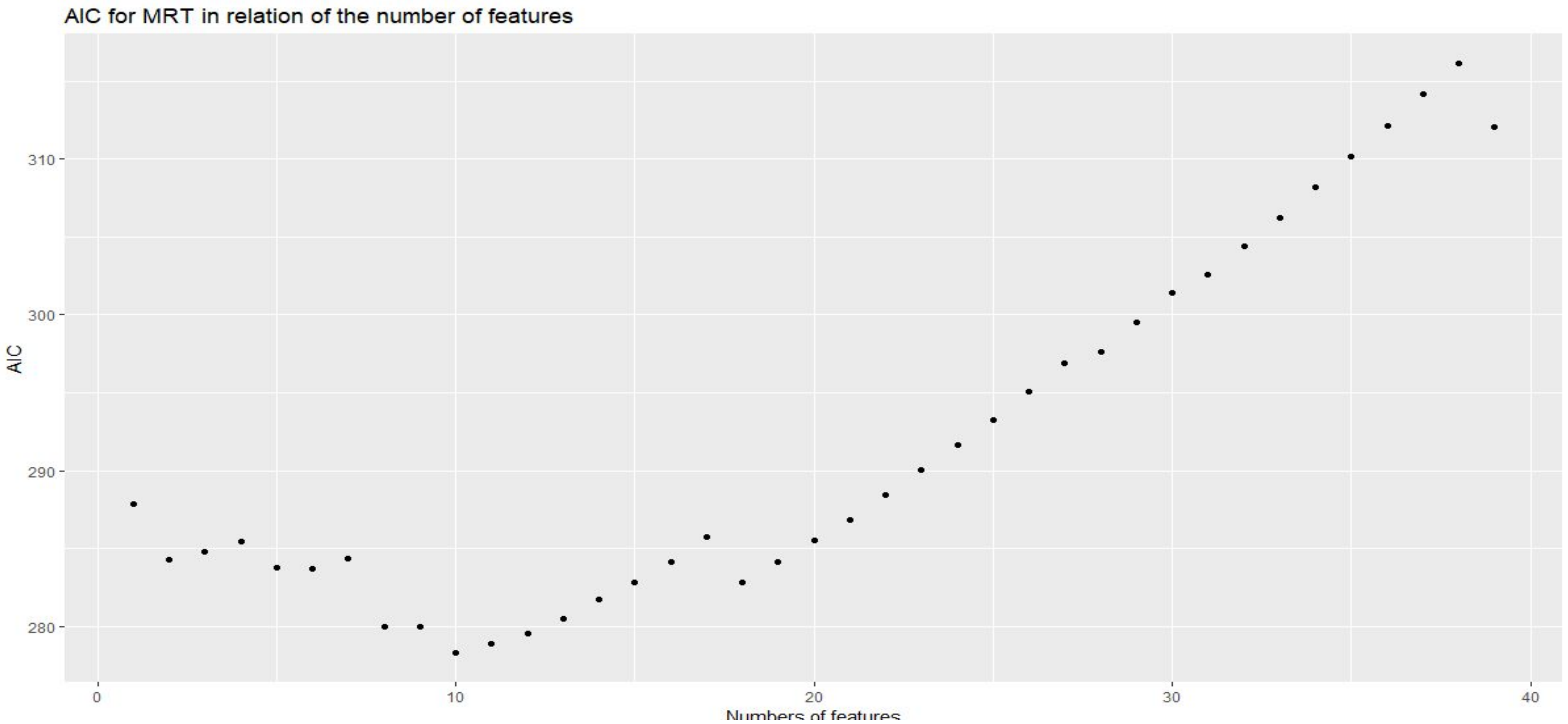


Model #5: Polynomial of degree two GLM

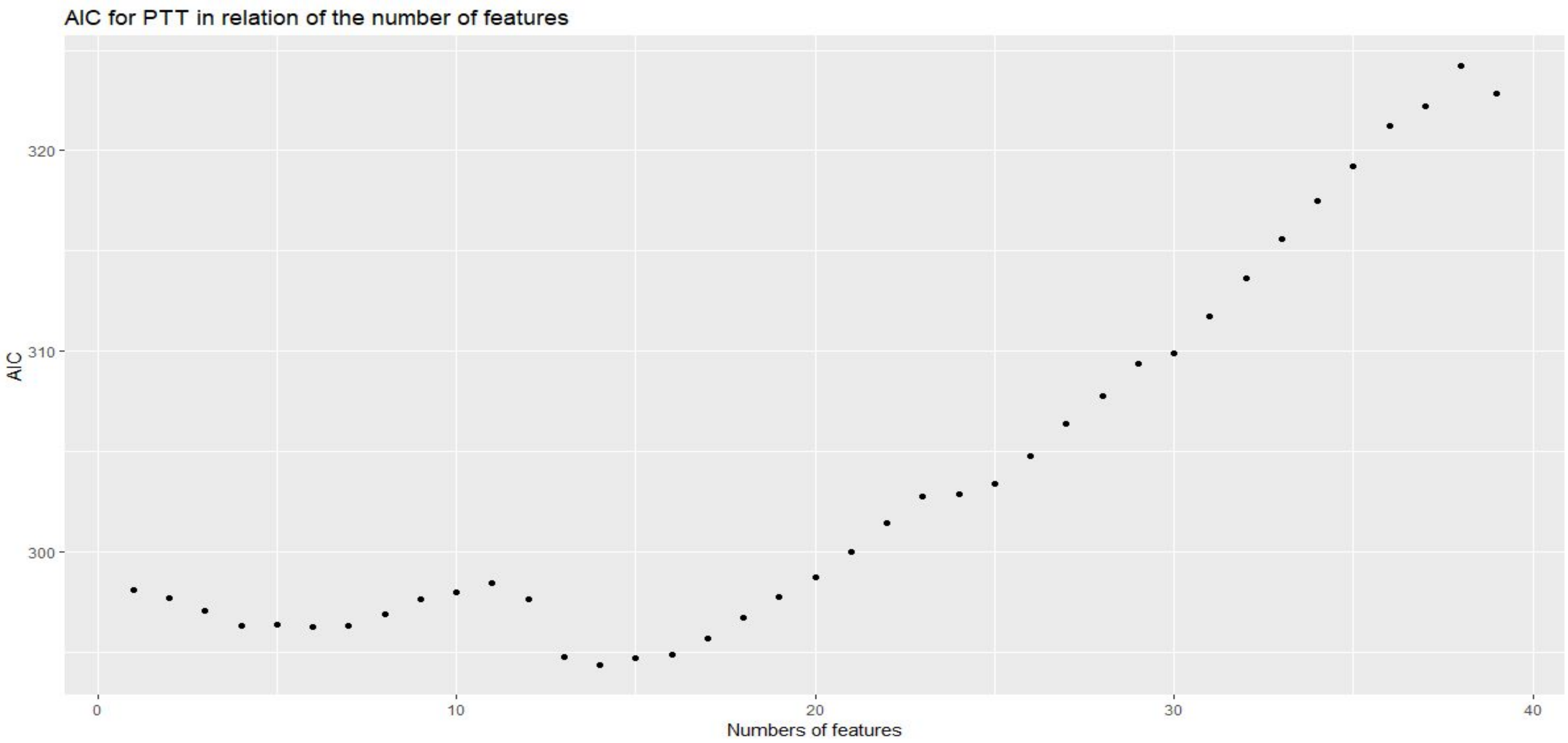
Having these mediocre results, we tried to run GLMs with a polynomial of degree two and the same procedure (AIC based) for features search.

The effect of using this new formula has given better result for MRT, by increasing the accuracy of 0.063, but worse for PTT, with a decreasing of accuracy of 0.013.

AIC by number of features in the case of MRT



AIC by number of features in the case of PTT

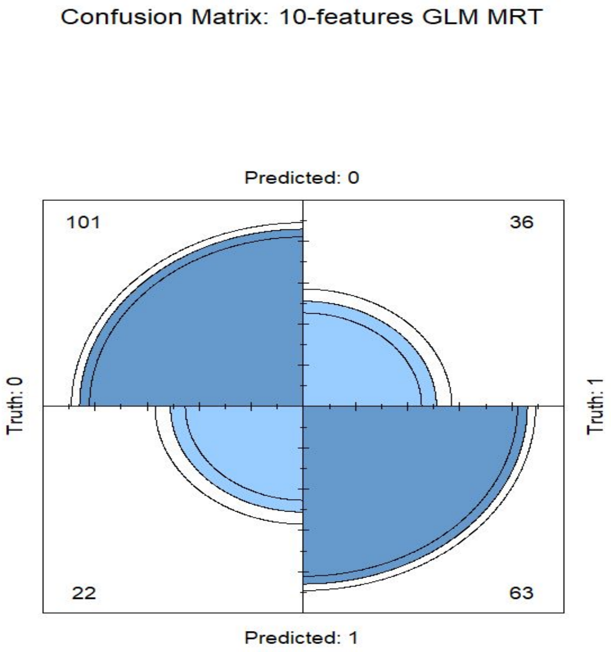


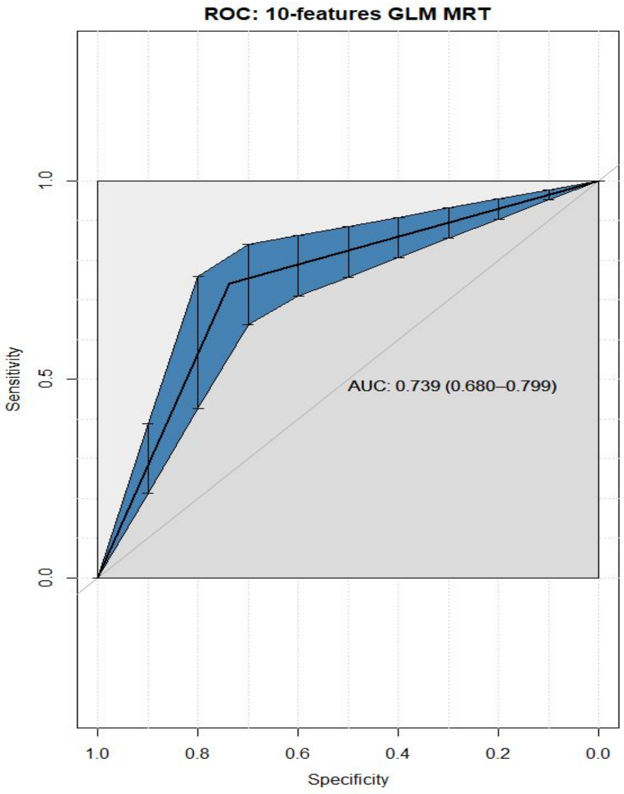
Result and best features of 2-poly GLM for MRT

1. Gender
2. Emotion Control**2
3. Cordiality
4. Cooperativity**2
5. Cooperativity
6. Culture Opening**2
7. Sense of Direction
8. Sense of Direction**2
9. Experience Opening
10. Experience Opening**2

1. Accuracy: 0.739
2. Sensitivity: 0.636
3. Specificity: 0.821
4. AUC: 0.739

Confusion Matrix and ROC Curve MRT



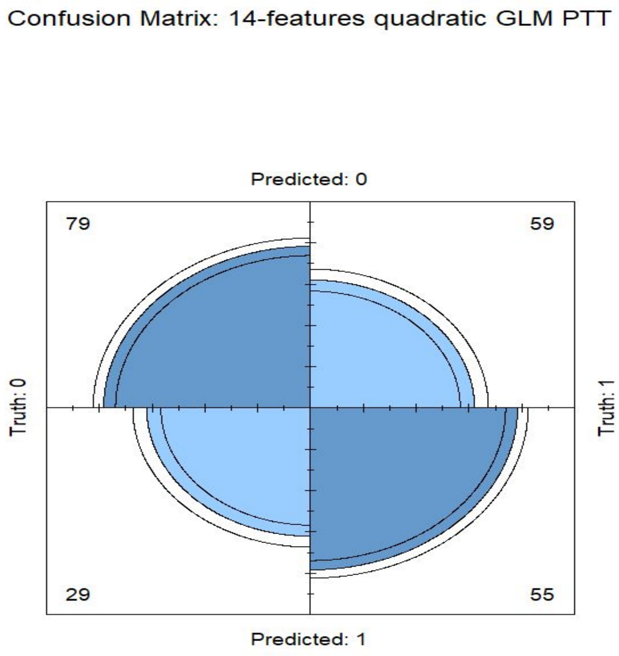


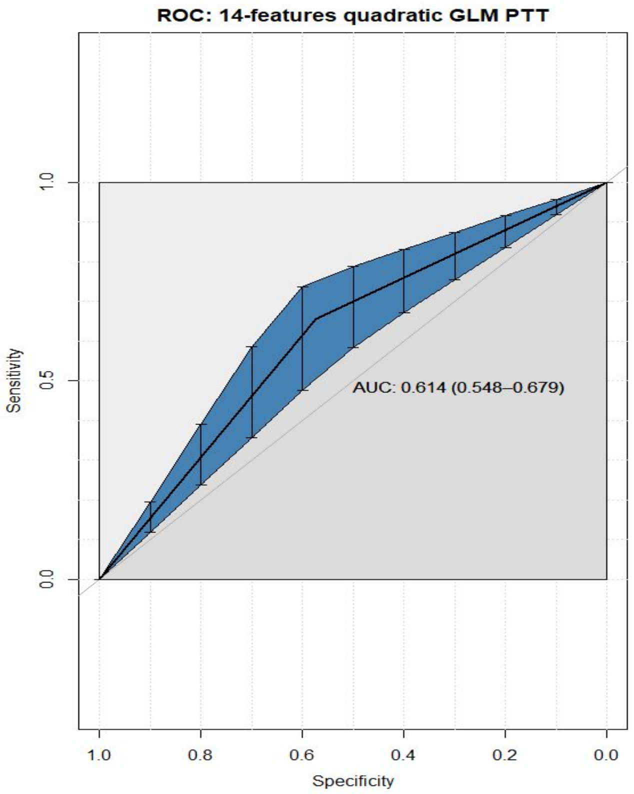
Result and best features of 2-poly GLM for PTT

1. Cardinal Points**2
2. Emotion Control**2
3. Emotion Control
4. Experience Opening
5. Perseverance
6. Map-Use**2
7. Cardinal Points
8. Map-Use
9. Culture Opening
10. Culture Opening**2
11. Cordiality**2
12. Cooperativity**2
13. Cooperativity
14. Experience Opening**2

1. Accuracy: 0.604
2. Sensitivity: 0.483
3. Specificity: 0.735
4. AUC: 0.614

Confusion Matrix and ROC Curve PTT





5.
Interpreting the data

Models comparison

A total of five models have been tried: K-NN, LDA, QDA, GLM (poly #1) and GLM (poly #2).

GLM models gave the best results, but still mediocre ones, especially in the case of PTT, meanwhile QDA seemed to be the worst one because its tendency to overfit too much.

Moreover MRT seemed easier to predict correctly for all the the five models

MRT accuracy

MRT models ranged from .653 to .739 of accuracy. In all the cases we have seen high accuracy when predicting the “low score” class (specificity) and a bad one to predict the “high score” class (sensitivity).

MRT	Features	Accuracy	Sensibility	Specificity	AUC
k-NN (k=18)	All	. 653	.434	.821	.656
LDA	6	.671	.526	.789	.670
QDA	2	.653	.526	.756	.649
GLM	2	.676	.535	.789	.674
GLM-poly	**10**	**.739**	**.636**	**.821**	**.739**

Table 4.1: *Results for the response variable MRT*

MRT features

Gender is for sure the most important features for Gender (present in all the model), followed by Emotion Control (present in 3 out of 4 model). Features present in two model were: Sense of Direction and Cardinal Points.

In the best performing model were also present some personality features

MRT	LDA	QDA	GLM	GLM-poly
1st	Gender	Gender	Gender	Gender
2nd	Emotion Control	Cardinal Points	Emotion Control	Emotion Control**2
3rd	Pulse Control	.	.	Cordiality
4th	QAS	.	.	Cooperativity**2
5th	Cardinal Points	.	.	Cooperativity
6th	Sense of Direction	.	.	Culture Opening**2
7th	.	.	.	Sense of Direction
8th	.	.	.	Sense of Direction**2
9th	.	.	.	Experience Opening
10th	.	.	.	Experience Opening**2

Table 4.2: *Features for the different MRT models*

PTT accuracy

PTT model never reached a good accuracy (0.559-0.617) having bad result in both sensitivity and specificity. The best model was a GLM with polynomial of degree one which reached 0.617 accuracy with a pretty balanced sensitivity/specificity (0.623, 0.611).

PTT	Features	Accuracy	Sensibility	Specificity	AUC
k-NN (k=18)	All	.563	.500	.629	.556
LDA	4	.595	.579	.611	.595
QDA	4	.559*	.526	.593	.560
GLM	**5**	**.617**	**.623**	.611	**.617**
GLM-poly	14	.604	.483	**.735**	.614

Table 4.3: *Results for the response variable PTT*

PTT features

If Gender was the main feature for MRT, for PTT the undiscussed most important feature was Cardinal Points, present in all the models. Gender, Map use and Verbal Indication were important for LDA and QDA, while Emotion Control and Experience and Culture Opening were important for GLM’s

PTT	LDA	QDA	GLM	GLM-poly
1st	Gender	Gender	Cardinal Points	Cardinal Points**2
2nd	Map Use	Map Use	Emotion Control	Emotion Control**2
3rd	Verbal Indication	Verbal Indication	Experience Opening	Emotion Control
4th	Cardinal Points	Cardinal Points	Perseverance	Experience Opening
5th	.	.	Culture Opening	Perseverance
6th	.	.	.	Map Use**2
7th	.	.	.	Cardinal Points
8th	.	.	.	Map Use
9th	.	.	.	Culture Opening
10th	.	.	.	Culture Opening**2
11th	.	.	.	Cordiality**2
12th	.	.	.	Cooperativity**2
13th	.	.	.	Cooperativity
14th	.	.	.	Experience Opening**2

Table 4.4: *Features for the different PTT models*

Final Discussion: not what we expected

MRT and PTT seemed to be not so easy to predict (PTT more than MRT) with GLM being the best performing model for both.

What is relevant and surprising from this work is not really the low accuracy of the model but the features which were relevant.

In fact, it’s pretty natural to think that the most relevant features would have been the ones coming from the self-assessment wayfinding inclinations questionnaire, since they ask aspects all linked to the field of spatial cognition and spacial behaviours, but this is not what happened since only Cardinal Points seemed to relevant (a lot for PTT, a bit for MRT), increasing the odd ratio of being in the “High score” class.

Final Discussion: Emotion Control

What came out from this analysis is a prominent role of Gender for MRT, Cardinal Points for PTT and Emotion Control for both the features.

While Gender differences (female performing worse than male) have been found in spatial cognition, the effect of Emotion Control has never been found.

This results suggest that higher Emotion Control increase the probability to have higher score in both the test.

Final Discussion: MRT and PTT, test under pressure

The effect of Emotion Control in both the two tasks can be explained looking at the condition of the subjects when taking the tasks. In fact, both PTT and MRT have a very limited time to be completed (no more than 5 minutes) and can be considered to be under pressure tasks.

This kind of condition could put in hurry the subjects, testing her/his ability to remain focus and calm during the task and this could easily explain the effect of Emotion Control.

Moreover, this could also explain the really relevant effect of Gender since there is a negative correlation (-0.163) between Gender and Emotion Control.

Final Discussion: PTT and MRT could be wrong

From this results we can hypothesize that PTT and MRT could be not very suitable to catch the spatial ability of a person since:

1. They are really not so strongly correlated (0.431). When they should if they are assessing the same concept.
2. Wayfinding inclinations that should explain very well the spatial cognition of a person are not useful to predict the class scores of this two task
3. Emotion Control suggests that these two tasks rather than testing only the spatial ability of a subject, test also her/his ability to perform under pressure.

End