**Introduction**

Machine learning, which is a sub-branch of artificial intelligence, is developing day by day and as it develops, new products or solutions based on machines emerge. Artificial intelligence, which plays a major role in the industrial sector, increases the speed and quality of production in all areas with the algorithms and architectures used. On the other hand, artificial intelligence and machine learning algorithms are used in almost every scientific field on the academic side.

Machine learning is a sub-discipline of artificial intelligence, which is basically inspired by the way humans learn, training electronic devices with algorithms and architectures, and as a result, learning like humans.

**Literature Review**

**Material and Methods**

In this project we used some of Machine Learning algorithms for predict better result.

While I was developing project, I used some technic for preprocessing. First of all, I used descriptive statistics for detect missing values or outliers values. In this step realized that age feature has some outliers. After that I used Label Encoder because I wanted to mark outliers value and convert it to 0. In my some other features has outliers values as well so I changed all of them respectively. As soon as finish my this part, I printed out null values and I saw that only in major feature has 1158 null values that is why I could declare 0 to them. After thşs step, I used visualization functions for see my features relationship. Every part of plots was looking well so my preprocessing phase has done. After preprocessing step, I used train\_test\_split function for split values. I used %30 test size and random state.

In the Algorithm development phase, I used 4 algorithms and their names are, Linear Regression, Decision Tree Regression, Random Forest Regression, and K Nearest Neighbours.

**Dataset Information**

Given data set has 2495 values and each value has 72 features.

Q1 - Q15

* The time spent answering each question was also recorded, and are stored in variables

E1 - E15

* The other following time elapses were also recorded:

introelapse The time spent on the introduction/landing page (in seconds)  
testelapse The time spent on the GCBS questions  
surveyelapse The time spent answering the rest of the demographic and survey questions

TIPI1 Extraverted, enthusiastic.  
TIPI2 Critical, quarrelsome.  
TIPI3 Dependable, self-disciplined.  
TIPI4 Anxious, easily upset.  
TIPI5 Open to new experiences, complex.  
TIPI6 Reserved, quiet.  
TIPI7 Sympathetic, warm.  
TIPI8 Disorganized, careless.  
TIPI9 Calm, emotionally stable.  
TIPI10 Conventional, uncreative.

The TIPI items were rated "I see myself as:" \_ such that

1 = Disagree strongly  
2 = Disagree moderately  
3 = Disagree a little  
4 = Neither agree nor disagree  
5 = Agree a little  
6 = Agree moderately  
7 = Agree strongly

The following items were presented as a check-list and subjects were instructed "In the grid below, check all the words whose definitions you are sure you know":

VCL1 boat  
VCL2 incoherent  
VCL3 pallid  
VCL4 robot  
VCL5 audible  
VCL6 cuivocal  
VCL7 paucity  
VCL8 epistemology  
VCL9 florted  
VCL10 decide  
VCL11 pastiche  
VCL12 verdid  
VCL13 abysmal  
VCL14 lucid  
VCL15 betray  
VCL16 funny

**education** "How much education have you completed?", 1=Less than high school, 2=High school, 3=University degree, 4=Graduate degree  
**urban** "What type of area did you live when you were a child?", 1=Rural (country side), 2=Suburban, 3=Urban (town, city)  
**gender** "What is your gender?", 1=Male, 2=Female, 3=Other  
**engnat** "Is English your native language?", 1=Yes, 2=No  
**age** "How many years old are you?"  
**hand** "What hand do you use to write with?", 1=Right, 2=Left, 3=Both  
**religion** "What is your religion?", 1=Agnostic, 2=Atheist, 3=Buddhist, 4=Christian (Catholic), 5=Christian (Mormon), 6=Christian (Protestant), 7=Christian (Other), 8=Hindu, 9=Jewish, 10=Muslim, 11=Sikh, 12=Other  
**orientation** "What is your sexual orientation?", 1=Heterosexual, 2=Bisexual, 3=Homosexual, 4=Asexual, 5=Other  
**race** "What is your race?", 1=Asian, 2=Arab, 3=Black, 4=Indigenous Australian, Native American or White\*\*\*, 5=Other  
\*voted\* "Have you voted in a national election in the past year?", 1=Yes, 2=No  
**married** "What is your marital status?", 1=Never married, 2=Currently married, 3=Previously married  
**familysize** "Including you, how many children did your mother have?"  
**major** "If you attended a university, what was your major (e.g. "psychology", "English", "civil engineering")?"

**Experimental Result**

In Liner Regression I used fit function and my model score is equal to 0.46898818603986725, Mean Squared Error(MSE) score is 147.28735791884492, Mean Absolute Error(MAE) score is 6.885437402291669 and also I have r2 score which is 0.16301728238274504

Secondly I used Decision Tree Regression, I reached predict values and model score, my model score is equal to 0.7262760087875616, Mean Squared Error(MSE) score is 163.00667556742323, Mean Absolute Error(MAE) score is 8.51001335113484 and also I have r2 score which is 0.07368987919960612

Thirdly I used K-Nearest Neighbors, I reached predict values and model score, my model score is equal to 0.39879759519038077, Mean Squared Error(MSE) score is 267.6502002670227, Mean Absolute Error(MAE) score is 11.36715620827770 and also I have r2 score which is -0.5209627978643561

Lastly I used Random Forest Regression. I had model score value which is 0.8299419659391916. The best algoritm is Random Forest Regressor for this dataset. it might be better if I usesome bigger n\_estimator values. I also had MSE score which is 75.48851281708944, MAE score 6.2589719626168225 and r2 score 0.5710250934004397.

**Conclusion**

In here I learnt how to preprocessed, assigned, visualized, and used regression Algorithms on it. When I was working on Project I learnt some of new value type for example scipy.stats or sklearn.metrics and that methods helped me a lot in that project.