Predicting Student Satisfaction Using Machine Learning

Student Name: M.C.M Julia

Supervisor Name: R. Crosby

Submitted as part of the degree of BSc Computer Science to the

Board of Examiners in the Department of Computer Sciences, Durham University

Abstract — This paper implements and evaluates the highest performing machine learning (ML) algorithms in determining student satisfaction in higher education, based on student demographic characteristics such as sex, age, domicile, ethnicity, and disability status. Using data from the National Student Survey (NSS) 2023, the study evaluates several ML algorithms using evaluation metrics and diagnostic plots. The evaluated ML algorithms are Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Linear Regression (LR), and Naïve Bayes (NB). The study identifies Random Forest as the highest performing algorithm for the prediction of student satisfaction. The study also evaluates the benefit and relevance of a display interface, a website, for students to interact with and predict their own satisfaction based on their characteristics. Despite limitations in sample size and local functionality, the feedback from students suggests positive benefits from such an interface. Further research needs to be conducted to fully establish the relevance of the interface. Ethical considerations regarding ML algorithmic accountability and fairness are highlighted, suggesting avenues for future research to refine interface design and address ethical concerns.

Index Terms— Interfaces, Machine Learning, Performance evaluation, Relevance feedback

1 Introduction

1.1 Background

In 2023, within the United Kingdom (UK) there were 757,000 applications for full-time undergraduate placements alone through the Universities and Colleges Admissions Service (UCAS) [1]. For applicant's student satisfaction has an impact on student choice although the degree of impact is often determined by its influence on league table rankings [2]. Within higher education student satisfaction emerges as a pivotal metric. The National Student Survey (NSS) assess student satisfaction around the UK states it gathers "students' opinions on the quality of their courses" [3] suggesting student satisfaction serves as a barometer for the quality of education, reflecting not only the effectiveness of teaching methods and institutional resources but also the overall support structure provided to students. Studies also imply that student satisfaction is influenced by many varying factors including demographic characteristics [4]. Therefore, possibly suggesting that a specifically tailored satisfaction rate to the student's demographic characteristics may be more relevant than the NSS data to the student and potentially better aid the student in their higher education choice.

In 2023 the NSS received over 339,000 responses with a response rate of 71.5 per cent [3]. This volume of data presents both opportunity and barriers, as for humans analysing such volume would provide difficulty. However, this is providing an opportunity for computer based predictive techniques. Machine Learning (ML) involves the computer analysing large amounts of data to find patterns and rules and then using those to characterise new data, giving the ability for a computer to learn from experience [5].

The application of ML for prediction within higher education has been continuously developed and researched over the years. However, research has been focused in predicting student performance [6], [7], [8], [9], [10], [11]. In recent years predicting student satisfaction has emerged as a critical endeavour with far-reaching implications for academic institutions [12], [13], policymakers [14], and students themselves [12], [15]. Research for student satisfaction predictions has been primarily focused on online learning [16], [17], [18], [19].

High levels of student satisfaction have been linked to increased academic engagement, improved academic performance, and overall well-being [20]. The significance of predicting student satisfaction extends beyond academic circles, with broader societal implications as higher satisfaction leads to higher retention rates contributing to a more educated and skilled workforce [21]. This is important as a higher educated society contributes to a successful society by benefiting a productive, increasing employment opportunities, greater social cohesion as described in [22].

The integration of ML algorithms into the prediction of student satisfaction represents a promising avenue for research and practice. By leveraging advanced computational techniques, this study can analyse vast datasets encompassing diverse student characteristics to accurately predict student satisfactions.

1.2 Research Questions

The research questions were formulated with the aim of bringing new knowledge into the field of predicting student satisfaction using machine learning and potentially create research that can be used and built on in the future to help prospective students in their choice. This project aims to answer the following two research questions (RQ):

RQ 1 – *What is the highest performing machine learning algorithm, according to evaluation metrics and diagnostic plots, to predict student satisfaction based on student characteristics?*

RQ 2 – *Based on student feedback, how beneficial and relevant is a display interface utilising the machine learning algorithms to predict student satisfaction based on characteristic to real life students?*

1.3 Objectives

The objectives of this study are split into four.

Objective 1: Use as predictive features at least three of the following characteristics to help predict student satisfaction: sex, age, domicile, ethnicity, and disability.

Objective 2: Use at least four different satisfaction themes from the seven NSS themes.

Objective 3: Implement and evaluate at least three of the following machine learning algorithms: Artificial Neural Network, K-Nearest Neighbour, Linear Regression, Naïve Bayes, Random Forest, and Support Vector Machine.

Objective 4: Create and implement an interactive display interface that will allow student to input their characteristics and it will output the student's satisfaction.

These objective contribute to the research question as with the completion of objective 1 and 3 the study should be able to produce an answer for RQ1. With the completion of objective 2 and 4 the display interface will built appropriately for students to evaluate its benefit and therefore aiding in answering RQ2.

2 Related Works

2.1 Student Satisfaction

Across research and literature student satisfaction has been defined in multiple examples. A generally accepted definition is "a short-term attitude resulting from an evaluation of student's educational experience" [2]. Clemes, Gan and Kao [4] further defined student satisfaction as the overall satisfaction with academic university experiences and service quality perceived by university students.

Student satisfaction has been observed as important. Kanwar and Sanjeeva [12] highlighted student satisfactions a crucial role in quality improvement in Higher Education (HE) as it provides valuable feedback and insights into student's experience, preferences, and needs. Stating that "students are the most important stakeholders of any education institution" [12] and therefore their satisfaction may lead to further contribution to the institution. This suggests a possible reasoning of why student satisfaction may be important to HE establishments. Student satisfaction is deemed important to students as it links to their quality of education, personal development, academic success [21], future opportunities, student's well-being, and academic engagement [12].

Factors influencing student satisfaction are vast. Kanwar and Sanjeeva [12] suggest that the factors can be split into two groups: personal and institutional factors. Personal factors could include examples such as personality traits, age, gender, or employment. Whilst institutional factors may include examples such as teaching quality, assessment and feedback or course content. Table 1 represents a compilation of factors and the study reference in which they have been identified as factors influencing student satisfaction.

The most frequent influential factors identified within the studies were teaching quality, assessment and feedback, accessibility, student demographic characteristics and course content. The factors identified were sometimes biased based on the context of the studies and would therefore possibly not have the same impact and influence within this study. For example, internet speed was identified as an influential factor in multiple studies however these studies were evaluating student satisfaction in an online learning environment. Therefore, although still a potential factor within an in-person environment the factor may not be as influential. Table 1 provides the understanding of the various influential factors. There is an identified influence of demographic characteristics within previous literature, but it has only been identified a few times suggesting an importance of this study to further explore this area.

TABLE 1
FACTORS INFLUENCING STUDENT SATISFACTION BASED ON LITERATURE

Factors	Literature Reference	Frequency
Teaching Quality	[4], [13], [16], [17], [23], [24], [25], [26], [18]	9
Assessment and feedback	[13], [23], [17], [24], [18]	5
Accessibility	[4], [16], [17], [24], [18]	5
Student Demographic Characteristics	[4], [23], [24], [15]	4
Course content	[4], [25], [18], [17]	4
Facilities	[4], [25], [26]	3
Internet speed	[16], [18], [17]	3
Organisational and Management aspect	[13], [4], [25]	3
Learning resource	[13], [24], [16]	3

Personal development opportunities	[13], [4]	2
Motivational variables	[21], [24]	2
Academic development	[4], [25]	2
Social Variables	[4], [26]	2
Student-instructor formal interaction	[15], [25]	2
Personality Traits	[4][19]	2
Student engagement	[13]	1
Cognitive and Achievement Variables	[21]	1
Institutional Variables	[23]	1
Workload	[24]	1
Classroom Environment	[25]	1
Accommodation	[26]	1
Student–student informal interaction	[15]	1
Student–student formal interaction	[15]	1

Within the NSS data the factors that were identified to influence and determine student satisfaction were: teaching quality, assessment and feedback, organizational and management aspects, learning resources, personal development opportunities, student engagement [13]. The ability of the NSS data to encompass a wide variety of determinants to students' satisfaction is important for this study as it allows for the incorporation of numerous data points and features making it an ideal dataset for machine learning analysis. Further suggested by the conclusion that "the NSS contains useful and detectable patterns of student behaviour" [27]. Moreover, its comprehensive coverage enables to predict specific themes of satisfaction and not just overall satisfaction which will allow students to understand the specific areas they may be satisfied with.

2.2 Student Performance

Student performance refers to the academic achievements and progress of students in higher education institutions. It encompasses various aspects such as grades, attendance, progress, and overall success in academic activities [6].

There is a consensus that student performance and student satisfaction are closely related and impact one another. The research conducted by Wach et al. [21] suggested a positive correlation between academic achievement and satisfaction with the conditions of the academic program, referring to the environment and circumstances that students experience at their university. Skrbinjek and Dermol [14] have also suggested that when students are satisfied with specific aspects including the quality of teaching, it can potentially lead to an improved academic performance. Soo [28] further detailed that the relationship between student satisfaction and student performance in UK universities varied based on whether the university was established pre or post the Further and Higher Education Act 1992. Student satisfaction was positively related to degree performance in post-92 universities but for universities pre-92 student satisfaction was negatively related to degree performance [28]. The NSS collects data from all UK universities, and nearly half of these are post 1992. Therefore, it may be important for the student to understand the difference when predicting its student satisfaction to understand further the implications of the prediction.

Research found that factors impacting student performance include demographic factors, academic, family, or personal, and internal assessment attributes with specifically attendance, gender, age and nationality being crucial to student performance [6].

With many studies conducted on predicting student performance using machine learning ([6], [7], [8] [9], [10], [11], [29]) and the close relationship established between student performance and student satisfaction there would be an interest in seeing whether the results and drawn conclusion in the studies are replicated with the prediction of student satisfaction.

2.3 Student Experience

Student experience refers to the totality of student's perceptions and interactions with different educational services, staff, and peers while they are students [30], although, its meaning often varies to fit the specific purposes of studies. A direct relationship between student experience and student satisfaction has been established within research with both student satisfaction and student performance being a subset of student experience as the quality of a student's experience directly impact both [31]. Although this research lacks in diversity within its data and is old the established relationship remains accepted within other more recent studies [30].

2.4 Machine Learning Algorithms

The use of machine learning to predict student satisfaction has mostly been explored within the context of online learning. The use of machine learning to predict student performance is a much more studied area. From the literature examined, common machine learning algorithms for predicting student satisfaction and performance include decision tree (DT), random forest (RF), K-nearest neighbour (KNN), support vector machines (SVM), artificial neural network (ANN), linear regression (LR), naïve bayes (NB). Table 2 highlights the machine learning algorithm and the study reference(s) they have been observed in.

TABLE 2
MACHINE LEARNING ALGORITHMS EXAMINED IN LITERATURE ABOUT STUDENT SATISFACTION AND PERFORMANCE

Factors	Literature Reference	Frequency
SVM	[6],[16], [29], [8], [10], [11], [18]	7
ANN	[6], [7], [17], [8], [11]	5
KNN	[6], [16], [29], [9], [18]	5
NB	[6], [29], [7], [8], [18]	5
DT	[6], [14], [7], [8]	4
RF	[16], [8], [9], [10]	4
LR	[29], [7], [18]	3
Gradient Boosting Machine / Trees (GBM/T)	[10], [18]	2
LightGBM	[16]	1
Multi-Layer Perceptron Regression (MLPR)	[16]	1

Elastic Net (ENet)	[16]	1
Sequential Forward Selection (SFS)	[9]	1
Pruned Tree Classifier (PT)	[9]	1
Teaching Learning Based Optimisation (TLBO)	[11]	1
Latent Dirichlet Allocation (LDA)	[29]	1
Classification and Regression Trees (CART)	[29]	1

The common algorithms identified and indicated from table 2 suggest a good basis of algorithms to explore and potentially use within this study.

2.4.1 Definitions

To be able to properly evaluate ML algorithms within previous research and be able to appropriately implement them it is important to define the algorithms and understand how they work.

DT utilizes a tree like structure to represent decisions and their potential outcomes, including chance events, costs, and utility [14].

RF utilises an ensemble of decision trees, selecting the most frequent class through voting and aggregating tree results [16]. This offers efficient predictions with minimal parameter tuning making it applicable to diverse populations and suitable for high-dimensional problems.

KNN is a straightforward nonparametric technique that selects the closest K training data points based on Euclidean distance to make predictions by averaging their target output values [16]. Therefore, it offers simplicity and ease of optimisation.

SVM finds the optimal hyperplane that best separates the data into different classes. This hyperplane is the one that maximizes the margin, or the distance between the hyperplane and the nearest data points from each class [29]. Support vector regression (SVR) was adapted from SVM to handle multivariate regression by constructing hyperplanes in high-dimensional space [32]. Therefore, providing a solution for nonlinear separable problems. This is important because many real-world datasets are inherently nonlinear and cannot be effectively modelled using linear techniques alone. By allowing for nonlinear relationships between the input variables and the target variable, support vector regression (SVR) enables more accurate modelling of complex data patterns. This capability is essential for addressing the diverse range of nonlinear relationships that may exist.

ANN is a computational system that emulates the human brain's function. It consists of interconnected artificial neuron or nodes, processing information through weighted connections. Structured in layers including input, hidden, and output layers, the network gathers, analyses and provides predictions or results [33]. Through supervised learning, ANNs adjust weights based on input and output data to minimize errors, enabling automatic prediction for new input data. Multilayer perceptron regression (MLPR) is a feedforward ANN it emulates the structure and asynchronous activity of the human nervous system, featuring input, hidden, and output layers of nodes capable of performing nonlinear activation functions to distinguish nonlinear data in supervised machine learning models [16].

LR is a statistical technique used to analyse the relationship between one or more independent variables and a continuous dependent variable. It aims to identify the best-fitting linear relationship between the variables, enabling prediction and inference based on the data [34].

NB is considered the simplest variation of the Bayesian network. This model assumes that every feature attribute is independent from the other attributes given the target attribute state. It uses a specific equation to calculate the target value based on the frequencies of attribute values in the training dataset [7].

2.4.2 Machine Learning Algorithms within Studies

Ho, Cheong, and Weldon [16] explored the use of six different machine learning models to predict student satisfaction with emergency remote learning (ERL). These six models were KNN, SVM, RF, light gradient boosting machine (LightGBM), MLPR, elastic net (ENet). After testing both ENet and MLPR, a type of ANN, outperformed the other machine learning algorithms. ENet was favourited over MLPR as MLPR limits further analysis on features' importance which was part of the goal of this study [16]. This would be disregarded within this study as it focuses on machine learning performance and not feature analysis suggesting a potentiality for ANN to perform well within this study. With recursive feature elimination performed, RF was consistently the better performing algorithm during training but then ENet would be better performing with the testing data. This may suggest that although RF could be a potentially high performing algorithm it also has the likelihood to overfit based on this study [16]. Therefore, during its implementation proper tools need to be put in place to help prevent overfitting.

The use of ANN to predict student satisfaction in E-learning was found to be 92.2% accurate in [17]. Further suggesting the high performance of ANN despite the small dataset it was based on which normally reduces ANN's predictive performance and efficiency [35]. The use of ANN was also observed within the prediction of student performance as in [7]. The study aimed to predict computer science student performance within a specific university. The study implements four machine learning algorithms, ANN, DT, logistic regression, NB. From these NB had the lowest accuracy at 66.52% and ANN had the highest accuracy at 77.04% indicating the high performance of ANN to predict student performance. With the previously established relationship between student performance and student satisfaction it could be implied that ANN would also be high performing to predict student satisfaction in this study.

Baashar et al. [6] performed a literature review over 30 studies about predicting student performance. The most frequent machine leaning algorithms used were KNN, ANN, NB, DT, SVM. Figure 1 below displays the distribution of machine learning approaches in student's performance. The continued use of these algorithms within studies suggests a consistent interest in implementing and testing these algorithms. This could therefore potentially imply that these algorithms are high performers and is the reason for their continuous use within studies. This therefore brings interest and reason into implementing and evaluating them within this study and context to see if their potential high performance is sustained or if results differ to expectations.

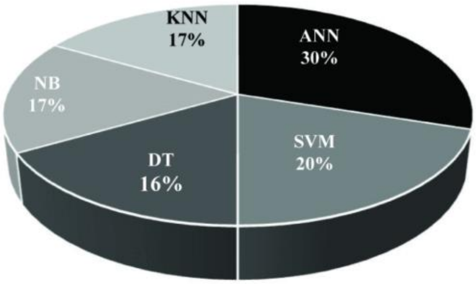


Fig 1. Distribution of machine learning approaches in student's performance [6].

The study highlighted that NB had the lowest maximum and minimum accuracy with 90% and 65.1% respectively [6]. ANN had produced the highest maximum and minimum accuracy at 98.3% and 73.5% respectively. Additionally, it had the smallest range suggesting that it most likely will predict with high accuracies consistently across different context and features. However, in [8] the literature review was done of 70 papers about predicting student performance and contrastingly to [6] it concluded that SVM performed with the highest accuracy. Although the conclusion is limited by the contexts the algorithm was applied in and the papers reviewed it offers a different high performing algorithm for the study. Therefore, in order to find the highest performing algorithm within this study it would be crucial to implement these two algorithms which have both been debated as the better performing algorithm in separate occasions. The authors also conclude that SVM, DT, NB, and RF have been well studied and generate good results implying they are worth testing and implementing for this study as they have been consistently performing well [8]. The author further details within its conclusion that ANN could be beneficially utilized for other types of predictions within the educational field further reasoning its inclusion within this study.

Gull et al. [29] further emphasised SVMs high performance ability as it aimed to predict students' final grades using machine learning. The study tested six algorithms: LR, linear discriminant analysis (LDA), KNN, classification and regression trees (CART), NB, SVM. LDA was found to have the highest accuracy with 81% however SVM was second highest with 80%. LR produced the lowest accuracy of 65% in this case. Although this suggest LR does not perform as highly as other algorithms its simplicity and insights into relationship could imply it to be a good use as a baseline algorithm to be used within this study [15]. This means it would be used not only to analyse the base relationship between independent and dependent variables but also to flag any potential anomalies in results from other algorithms by comparing them to the results of LR. Although ENet which is a version of LR had the highest performance in another study suggesting that LR does still have the potential to perform well [16].

DT, although a well-studied machine learning algorithm as depicted by [6], [8], its variant RF is considered as a better performing algorithm as it predicts with higher accuracy [36]. RF is a variant of DT and although it takes longer in time to carry out a task its higher accuracy prediction rate would be prioritised for this study. Gray and Perkins [9] had found that RF performed with the highest accuracy out of all the machine learning algorithms used, preferring it to random trees (RT) another variant to DT. Therefore, given that RF is a variant of DT and exhibits superior predictive performance compared to DT, it is superfluous for the study to assess DT when RF is substantially anticipated to surpass it.

Some studies such as [10],[11] preferred a hybridised model mixing two machine learning algorithms to help predict student performance. Both had better accuracy with the hybrid model than using the machine learning algorithms alone. Both decided to use SVM with [10] mixing it to RF and gradient boosting (GB) and [11] mixing it with ANN. However, this study aims to first identify the higher performing machine learning models on their own and hybridised models could be a potential future work built on top of this study.

Overall, according to the research it is within interest for the study to implement and evaluate RF, SVM and ANN as they have the potential to perform the highest. With further implementation and evaluation of LR, NB, and KNN as although evidenced to perform poorly they have also been evidenced to perform well and would therefore be interesting to involve them within this study too.

3 METHODOLOGY

3.1 Data Selection

The data used is the NSS 2023 by characteristics data [37]. The data is available on the Office for Student (OfS) website [38]. The NSS data was selected as it is publicly available and collects data from all UK universities. This offers a wide pool of data but additionally is an area of the world where prediction of student satisfaction or performance through machine learning has not been widely researched and therefore offers interest in doing the study.

The data is composed of 19 different characteristics, see Appendix 1 for the full list, each separated into their own dataset and each dataset has 25 attributes, see Appendix 2 for the full list.

3.2 Preprocessing

3.2.1 Data Exploration

The coding for the data exploration was done in Python as it is a common programming language and has the necessary libraries and toolkits for a simple and efficient data exploration. The data exploration was minimal as NSS organise and explain their data well. It consisted of describing the data so for each characteristic, for the categorical attributes describing the count, unique values, top value, and frequency. For the discrete attributes describing the mean, standard deviation, minimum value, 25th percentile, 50th percentile, 75th percentile, and the maximum value. In addition to this the datasets were checked for missing values. This allows to indicate any data cleaning and preprocessing steps that may need to be carried out. It also helps better understand the datasets and their components to better adjust machine learning algorithms but also help understand results later.

3.2.2 Feature Selection

There is two type of feature selection that needs to be carried out. The characteristics need to be selected and the appropriate attributes need to be selected, both were done manually. These were done manually as the data was nicely set up in a spreadsheet and therefore it was more appropriate to remove them and edit the spreadsheet. Additionally, the attributes were selected manually as not all attributes were appropriate and would benefit the ML algorithms. Such as the multiple benchmark attributes, these were irrelevant to predicting student satisfaction as they were linked to some data exploration the NSS had done and were not data directly collected from the survey.

The characteristics selected for the research were sex, age, domicile, ethnicity, and disability. These were selected as they had been identified within other studies as demographic factors that influence student satisfaction [21].

The attributes selected for each characteristic were, split, provider country, mode of study, level of study, subject, question number, responses, and positivity measure. These were selected as they were relevant and appropriate to predict student satisfaction. The split is the values for the given characteristic. The provider country includes options England, Northern Ireland, Scotland, Wales, UK. The mode of study includes options full-time, part-time, apprenticeship, all modes. The level of study includes options first degree, other undergraduate, undergraduate with postgraduate component, all undergraduates. The subject is the subject name. The question number is the number of the question or the theme number, the study does not use the question specific responses but instead looks at the seven satisfaction themes identified by the NSS and will aim to predict a student satisfaction for each. The responses are the full person equivalence number of responses to each question excluding those who answered, 'This does not apply to me'. The positivity measure is the proportion of responses that were positive in percentage essentially indicating student satisfaction. These attributes were selected as they were the only relevant attributes for the prediction of the positivity measure which is the student satisfaction. No further feature selection technique was then carried on the smaller attribute group as the feature numbers were already reduced and a lack of features would negatively impact machine learning algorithms [39]. Additionally, further reduction of the feature set would require using bootstrap aggregation to help counter the limitations of low dimensionality, which does not always consistently improve the model and in some cases can reduce accuracy as seen in [39].

3.2.3 Data Manual Preprocessing

From the data exploration it was observed that the data is clean and requires no modification in that matter however with the feature selection completed the irrelevant characteristics and attributes needed to be removed and the characteristics datasets needed to be separated into individual files to they can be processed individually.

3.2.4 Data Preprocessing

With six algorithms needing to be built data preprocessing steps were generalised for all to reduce bias in performance coming from different preprocessing steps and allow for a fair evaluation of performance. For these steps python was used, using the panda's library [40] and scikit-learn toolkit [41]. Pandas was used as it is said for data preparation to be "the best and most used Python library in this field" [42]. It offers a plethora of capabilities for handling various data formats and seamlessly collaborates with scikit learn [42]. Scikit-learn toolkit was used as it is said to be the "most comprehensive and open-sources machine learning package in Python" [43] and would therefore be adapted to this machine learning based study. Additionally, it has good documentation and unified bug fixing process on GitHub making it ideal if any issues were encountered whilst coding the algorithms up. The data is made into a dataframe and then it is split into three dataframes X, y, and r with y being the target column positivity measure, X is all the other attributes and r is the responses number which will be used as a class weight. The data is then split into training and testing sets using an 80-20 split as seen in [16]. This was done before encoding the data to avoid data leakage, meaning the model will not be influence by the testing data during preprocessing [39].

The features containing categorical values were then encoded using One-hot encoding (OHE) converting them to a binary numerical format which allows machine learning modelling [16]. To avoid overfitting and inflated results the training dataset was split into 10 independent sets of observations (K= 10) using K-fold cross validation method [16]. Standard scaler was used to further standardise the data for better model performance by ensuring that each feature contributes equally to the analysis of student satisfaction [44].

Class weighting is used to adjust for class imbalances as the data is collected in a way that it's a collection of types of responses so to avoid the model being biased towards a certain split of the characteristic that has the most responses. This is done by using the 'comput_sample_weight' tool from scikit-learn and it is computed based on class frequencies which is the number of responses, r. The computed class weights are utilized during model trainings to adjust the models accordingly in paying attention to minority classes too, improving predictive performance. Therefore, helping mitigate any imbalances within the data.

3.3 Training Machine Learning Algorithms

3.3.1 ANN

Initially to build the ANN model, an MLPR model had been built. However, due to issues with the MLPR package and compatibility, TensorFlow's Kera's API was used instead.

Within each iteration of the K-fold cross-validation loop, a neural network model is defined using TensorFlow. The model architecture consists of a single hidden layer with a specified number of neurons, hyperparameter Hidden Neurons, and an output layer. The hidden layer uses the rectified linear unit (ReLU) activation function which introduces non-linearity to the model and helps capture complex patterns in the data [45]. The output layer uses linear activation, suitable for regression tasks as it directly output continuous values.

After defining the model architecture, it is compiled. Compilation involves specifying the optimizer, loss function and optional metrics. In this case the Adam optimiser is chosen which is an adaptive learning rate optimization algorithm known for its effectiveness in training neural networks [46]. The loss function chosen is mean squared error (MSE) which is appropriate for regression tasks as it measured the average squared difference between predicted and true values [47].

To tune the model nested loops are used to iterate through different combinations of hyperparameter values, hidden neurons and learning rate, as seen in table 3. The hyperparameter learning rate controls the step size at which the model parameters are updated during the training process. The possible values are selected as they are the common values tested and are also studied to be optimal values [48]. Although the ANN trained within this study differed slightly from the one used within the research. The comparison of the model before hyperparameter tuning and after can be seen in the results section.

TABLE 3
HYPERPARAMETER FOR ANN

Hyperparameters	Possible Values		
Hidden Neurons	32	64	128
Learning Rate	0.001	0.01	0.1

This exhaustive search approach allows exploring various combinations of hyperparameters to find the optimal configuration. For each combination of hyperparameters the model is trained on the training fold of the data. This ensures that the model is fitted to a subset of the data while reserving the validation fold for evaluating its performance. After training, the model's performance is evaluated using MSE on the validation fold. MSE serves as the metric to assess how well the model generalizes to unseen data based on the chosen hyperparameters.

The model with the lowest MSE on the validation set is considered the best performing model among all hyperparameter combinations. This model is selected for further evaluation on the entire validation set from the fold. This is done to assess the fold's model ability and select the best performing model from all the folds.

Once the best model selected it is test and then evaluated using the evaluation metrics and diagnostic plots to inspect its performance as described in section 3.4.

3.3.2 KNN

The model was built using sci-kit learn's KNeighboursRegressor model. The hyperparameters are tuned using grid search with cross-validation [49]. Grid search is one of the most commonly used methods and can be considered brute-force and inefficient for any high-dimensionality hyperparameter configuration space. However, within this study the configuration space is minimal and therefore grid search is an ideal tuning method [50]. The study [50] concludes Bayesian Optimization Hyperband (BOHB) to be the recommended method for optimisation contrastingly this study [51] suggests Grid Search to be the better performing hyper optimisation method, both agree that Grid Search is not the most time efficient. This technique exhaustively searches through the specified hyperparameter space and evaluates each combination using cross-validation. The goal is to identify the optimal combination of hyperparameters that minimises the MSE on the validation set. By tuning the hyperparameters the model's performance is optimised to achieve the best possible results.

The highest performing model with the most optimised hyperparameters is selected to run on the testing data. The results are then evaluated using the evaluation metrics and visualisation methods.

The hyperparameters tuned were N-Neighbours and P. The values selected for these can be seen in table 4. N-Neighbours, number of neighbours to use by default, is the main hyperparameter with P being optional hyperparameters. P is a hyperparameter for power for the Minkowski metric where 1 is equivalent to using the Manhattan distance and 2 the default value uses the Euclidean distance. Although optional it is ideal to try both in the event that the use of the Manhattan distance optimises the algorithm.

TABLE 4
HYPERPARAMETER FOR KNN

Hyperparameters	Possible Values				
N-Neighbours	1	3	5	7	9
P	1		2		

The values selected for N-Neighbours were selected as the optimal range is 1-10 [51], with the default value in scikit-learn being 5, the values were increments of 2 up and down from the default within the range. It was done in increments of 2 to limit the grids dimensionality and keep the grid search efficient.

3.3.3 NB

The model was built using scikit-learn. The regression model of NB, Gaussian NB can only be used on discrete data. This was an issue as the target variable, positivity measure, was continuous. A ten-bin discretisation method was applied to the target data to be able to use NB [52]. This was done during the preprocessing stage and performing discretisation's methods are effective and do not negatively impact the results and model as seen in [52].

The hyperparameters were tuned using grid search to find the optimal value for 'var_smoothing', a smoothing parameter used to stabilise variance estimation in Gaussian NB. The selected values for this hyperparameter were: 1e-9, 1e-8, 1e-7, 1e-6, 1e-5. These values were selected as the range is an optimal and appropriate for hypertuning Gaussian NB [53]

After selecting and identifying the best model with the most optimised hyperparameters the model is ran on the testing data and evaluated using evaluation metrics and visualisation methods.

3.3.4 RF

The model was built using scikit-learns random forest regressor. The hyperparameters were tuned using grid search cross-validation. The hyperparameters included: 'n_estimators', 'max_depth', 'min_samples_split', 'min_samples_leaf'. The grid search is performed within each fold of the K-fold cross-validation loop. The values used for each hyperparameter are described in table 5. These values were selected as they were in a common range for hyperparameters of RF and in increments that would cover the majority of the range but keep an efficient dimensionality for the grid search [50].

TABLE 5
HYPERPARAMETER FOR RF

Hyperparameters	Possible Values			
n_estimators	50		100	200
max_depth	None	10	20	30
min_samples_split	2		5	10
min_samples_leaf	1		2	4

The trained models are evaluated using MSE and the best model and hyperparameters is then used on the testing data. This model is then evaluated using the evaluation metrics and visualisation methods.

3.3.5 SVM

The model was built using SVR tool in scikit-learn. The hyperparameters were tuned using grid search cross-validation and included: 'C', 'gamma', 'kernel'. The grid search was performed within each fold of the K-fold cross-validation loop. C is the penalizing factor that controls the trade-off between the model complexity and the training error [54]. The kernel is the function used to map the input data into a higher dimensional feature space where a linear relationship between the input variables and the target variable may be easier to capture [54]. Gamma is the tuning parameter that controls the width of the Gaussian kernel function used in the SVR model [54]. The possible values for each are described within table 6. These were ideal candidate values as identified in research [54].

TABLE 6
HYPERPARAMETER FOR SVM

Hyperparameters	Possible Values			
C	0.1	1	10	100
gamma	scale	auto	0.1	1
kernel	linear		rbf	

The best model obtained from the tuning was then trained on the entire fold and evaluated on the validation fold using MSE. The best model and hyperparameters obtained from cross-validation are evaluated on the unseen testing data. The evaluation metrics and visualisation methods are then conducted on it.

3.3.6 LR

The LR model was constructed using LR tool in scikit-learn. For LR hyperparameter tuning is a little different, scikit-learn has different classes which encapsulate a CV check for the inner parameter and does not require to specify the value for alpha, the LR hyperparameter [55]. The names of these classes are LassoCV, RidgeCV, ElasticNetCV. All three variation models were trained and tested on the validation set in order to identify the optimal model. Following the completion of all the folds the better performing model with its corresponding hyperparameters were evaluated on the unseen testing dataset.

3.4 Evaluation Metrics

To evaluate the algorithms evaluation metrics were used to assess the performance of the algorithms. These were MSE, mean absolute error (MAE), r-squared score (R^2). Visualisation methods were also used to further understand and see the model's performance were residual plot and prediction error plot. Matplotlib, a library in python, was used to help create these diagnostic plot.

MSE measures the average squared difference between the predicted values and the actual values [47]. It is calculated by averaging the squared residuals. A lower MSE indicates that the model's predictions are closer to the actual values implying a better performance and providing a measure of the model's accuracy. However, MSE penalizes larger errors more heavily due to its squared nature and therefore is more sensitive to outliers.

MAE measures the average absolute difference between the predicted values and the actual values, it is calculated by averaging the absolute residuals [47]. MAE represents the average magnitude of errors in the predictions and a lower MAE indicates better performance. MAE is less sensitive to outliers, and it provides a straightforward measure of the model's performance in terms of prediction accuracy.

R^2 is a statistical measure that represents the proportion of the variance in the dependent variable (target) that is explained by the independent variables (features) in the model [47]. It ranges from 0 to 1, with the higher values indicating better fit.

Residual plot displays the residuals which are the differences between the observed and predicted values, against the predicted values. Residual plots help assess the goodness of fit of a regression model and its validity. Points on a residual plot as to be randomly distributed for a model to be considered valid and any models with trends in the plot is not valid [56]. Ideally the residuals should have a high density of points close to the origin and a low density of points away from the origin with symmetry at the origin to suggest strong fit. A residual plot requires points to be normally distributed to agree with the assumption of a good fit model. This then portrays the model's suitability to predict student satisfaction [57]. An example of a residual plot that demonstrates a valid model can be visualised in figure 2.

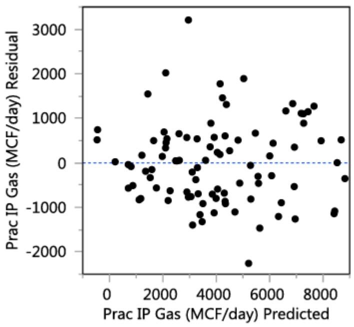


Fig 2. Residual plot demonstrating a valid model [58].

Prediction error plot compares the true values of the target variable against the predicted values generated by the model. The plot displays how well the model's predictions align with the actual values. Ideally the points should lie close to the diagonal line, indicating perfect prediction and deviations from the diagonal line indicates discrepancies between the predicted and actual values. The plot provides visual assessment of a model's accuracy and bias [58]. An example of a prediction error plot can be visualised in figure 3. This displays a well performing model with very minimal number of discrepancies.

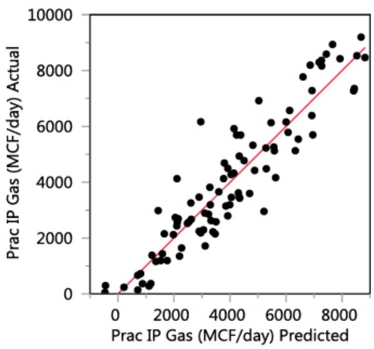


Fig 3. Prediction error plot with minimal discrepancies [58].

3.5 Display Interface

The display interface constructed was a website. It was selected to be a website because website can be easily accessed by all and can be used by future students easily to predict their satisfactions. Given the nascent nature of the study, only a prototype was built as it was found most appropriate until further feedback and evaluation was conducted. This means that the website could only run locally.

3.5.1 Front End

The front end was built using hypertext markup language (HTML), cascading style sheets (CSS) and JavaScript (JS). HTML is used for structuring the webpage's content. In the provided code it defines the layout of the webpage and CSS allows to style the webpage.

JavaScript is used to handle user interactions such as selecting options from dropdown menus and clicking buttons. Event listeners are added to respond to these interactions, triggering the appropriate actions such as when clicking the 'Predict Satisfaction' button in Figure 9. Asynchronous JavaScript and XML (AJAX) is used to make asynchronous requests to the backend without reloading the entire webpage. This allows for a smoother user experience as the page can update dynamically based on the server's response. JavaScript is also used to manipulate the Document Object Model (DOM), dynamically updating the webpage's content based on user actions and server responses. For example, hiding or showing elements in dropdown menus such as in Figure 6.

Figure 4 shows the first thing the user sees upon entering the website. The warning message is crucial as predicting something like student satisfaction can have an impact on the student's perception of their experience and performance at university. However, the model only considers a very limited number of factors that make up student satisfaction for the purpose of this study and is therefore not reflective of a student's actual satisfaction. This is important for any user who wishes to interact with the interface.

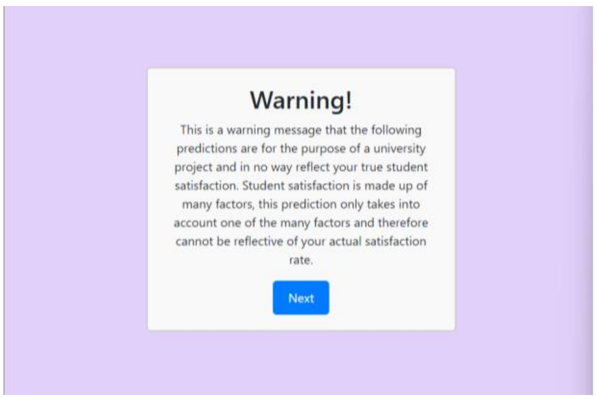


Fig 4. Warning message displayed to user after first loading the website.

Figures 5, 6, 7, 8, 9 display screenshots of the websites front end and various features the user interacts with.

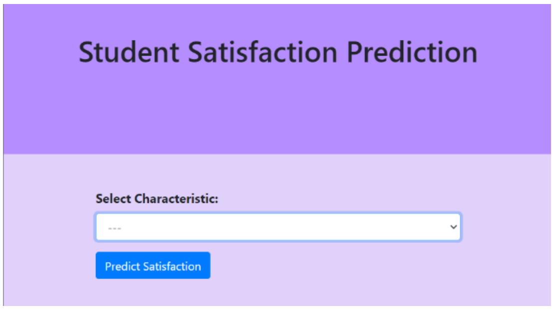


Fig 5. The next page after clicking the button 'Next'.

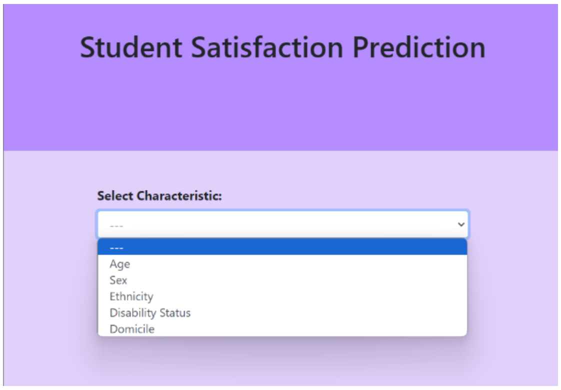


Fig 6. Dropdown menu with all the characteristic options

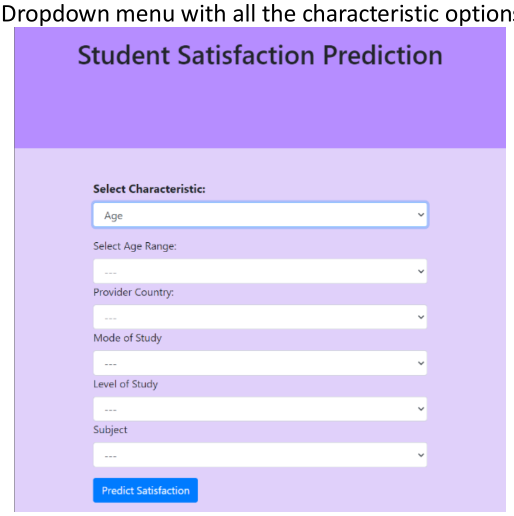


Fig 7. The website dynamically changing based on user option.

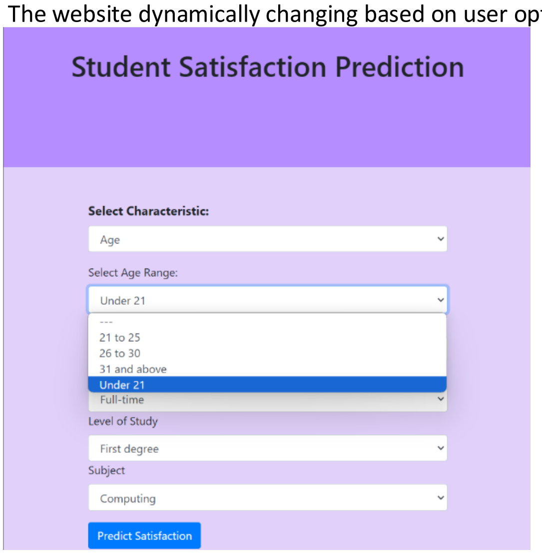


Fig 8. Dropdown menus with the options for the user to enter.

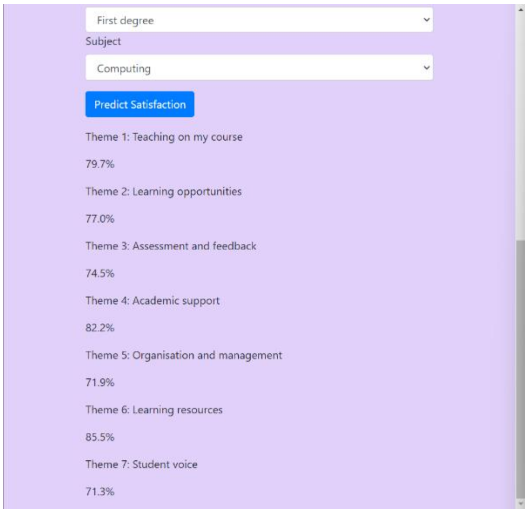


Fig 9. The student satisfaction prediction after the 'Predict Satisfaction' button is pressed.

Figure 9 contributes towards the completion of objective 2 as it illustrates the prediction of all seven NSS satisfaction themes. As well, figures 4-9 demonstrate the completion of objective 4 as an interactive display interface that will allow student to input their characteristics and output the student's satisfaction was created and implemented as evidenced by the figures.

3.5.2 Back End

For the back end the first step was to build the models that would run for each characteristic when the user wants to predict their satisfaction for it. To do this for each characteristics the models were compared to find the best performing model and then the parameters used for it was noted down so they could then be used to build the prediction model based on the model trained.

The prediction model is then trained and built in Flask (app.py). Flask is a lightweight web framework for python. It's used here for handling hypertext transfer protocol (HTTP) requests, routing, and serving web pages. Flask is chosen for its simplicity, flexibility, and ease of use in building web applications.

When the interface is first run the data is loaded into Pandas dataframe and the data is pre-processed as done in the models earlier. The '/predict' endpoint handles POST requests containing user data sent from the front end required for the prediction of user satisfaction. When the POST request is received the data is parsed and encoded, the satisfaction predictions are then made for each of the seven themes using the characteristics corresponding machine learning model. The Flask application is run using 'app.run()' to start the web server locally for development and testing.

3.5.3 Testing

Testing was conducted on the display interface. Both alpha and beta testing was performed. With alpha testing all the features, functions and prediction possibilities were tested which is when two issues were found in predicting features. The first issue was surrounding a missing feature for prediction. During the data preparation when the interface starts up one of the columns of the data had been accidentally deleted and therefore caused issues as the user was entering more features than expected. The second issue was with the OHE of categorical features. Only one one-hot encoder was used however this was causing issues if the user went back to predict another characteristic as the OHE had only transformed the split of that characteristic and would therefore not have fit the new split data and would not be able to transform the user data. The solution to this was to assign each model its own OHE that was fit during the training and when the user requested a prediction according to the characteristic the specific model and OHE was selected.

Beta testing was conducted using a feedback form. Through convenience sampling university students were selected to try and use the prototype for their own student predictions. Once finished they would fill in an anonymous feedback form which can be found in Appendix 3. This form evaluated the usability of the website and the practicality of the interface and the project, meaning users were asked whether this was a tool they wanted or find beneficial with the aim of getting results that would offer a potential answer to RQ2.

4 RESULTS

4.1 Machine Learning Results

The results of the implementation of the models on the testing data per characteristics can be found as follows. The averages in each table were built excluding the NB model results. The average represents the average model performance for that evaluation metric.

4.1.1 Sex

Table 7 bellow shows the results of the evaluation metrics on each model built for the sex characteristic. LR had the lowest performance with an MSE of 46.76 and an R-squared, value of 0.43. KNN and ANN exhibit similar performances with R-squared metrics of 0.6 and 0.62 respectively indicating a relatively better fit to the data. SVM and RF outperform the other models with RF being observed to have the highest performance as it has the best result in all three metrics and has an R-squared metric of 0.86 suggesting a strong predictive capability [47].

TABLE 7
EVALUATION METRICS RESULTS FOR MACHINE LEARNING ALGORITHMS FOR SEX

Model	Evaluation Metrics		
	MSE	MAE	R^2
LR	46.76	4.79	0.43
NB	5709.58	75.09	-68.25
KNN	33.16	3.82	0.6
RF	11.82	1.99	0.86
ANN	31.16	3.93	0.62
SVM	24.38	2.99	0.7
Average (without NB)	29.46	3.5	0.64

NB can be seen with extremely poor evaluation metric results. This indicates that something went wrong with the model. The validation sets MSE had averaged a 6.88, as seen in appendix 4 which has all the validation set results. NB thus, would have been the highest performing model. Therefore, the good performance during training and poor performance during testing suggests that there has been severe overfitting during the models training and it therefore is unable to perform in the testing phase. The overfitting is only seen with the NB model suggesting that it may be caused by the discretisation as that's the only training and preprocessing that is different to the other models. As other models are performing well NB is ruled out but for future studies different discretisation technique should be used as suggested in [52].

According to Table 7 the answer to RQ1 would be that RF is the highest performing ML algorithm to predict student satisfaction. RF being the highest performing was expected according to study [9]. RF also did not overfit during training unlike it was suggested it would by research [16]. This can be observed from its average validation set MSE result being 14.54. This is a slightly lower average performance than its performance on the test set and therefore the methodology to build the model was appropriate to prevent overfitting.

4.1.2 Age

Table 8 shows the results of the evaluation metrics on each model built for the age characteristics. Similarly to the models built for the sex characteristic both RF and SVM were the higher performing algorithms with RF being the highest performing algorithm. Both algorithms have performed better than with the Sex characteristic.

Similarly, NB's results are very poor and has been overfitting. Further suggesting the issue might be with the discretisation's as the issue stays the same when the model is trained on different data.

The average model evaluation metric results are also better than for any other characteristic. The superior performance of machine learning algorithms in predicting the age characteristics may suggest that age has a more significant influence on student satisfaction. Although previous research had suggested sex was the most influential demographic characteristic [6]. The difference in conclusion may be due to the NSS data and that within the context of the NSS data this study determines that the age characteristic may be the more influential demographic characteristic.

TABLE 8
EVALUATION METRICS RESULTS FOR MACHINE LEARNING ALGORITHMS FOR AGE

Model	Evaluation Metrics		
	MSE	MAE	R^2
LR	41.7	4.67	0.52
NB	5732.33	75.24	-64.35
KNN	30.76	4.07	0.65
RF	11.74	2.09	0.87
ANN	26.54	3.79	0.7
SVM	15.16	2.26	0.82

Average (without NB)	25.18	3.38	0.71

4.1.3 Domicile

Table 9 displays the results of the evaluations metrics on each model built for the domicile characteristic. A pattern is starting to be established as once more RF and SVM are the better performing algorithms with RF performing the highest and NB is very poorly performing.

TABLE 9
EVALUATION METRICS RESULTS FOR MACHINE LEARNING ALGORITHMS FOR DOMICILE

Model	Evaluation Metrics		
	MSE	MAE	R^2
LR	39.51	4.48	0.44
NB	5827.28	75.96	-82.31
KNN	32.16	3.74	0.54
RF	10.48	1.84	0.85
ANN	27.83	3.66	0.6
SVM	23.62	2.9	0.66
Average (without NB)	26.72	3.32	0.62

4.1.4 Ethnicity

Table 10 displays the results of the evaluation metrics for the models trained to the ethnicity characteristics. Ethnicity models have similar results and patterns than the other characteristics.

TABLE 10
EVALUATION METRICS RESULTS FOR MACHINE LEARNING ALGORITHMS FOR ETHNICITY

Model	Evaluation Metrics		
	MSE	MAE	R^2
LR	42.87	4.67	0.44
NB	5613.1	74.51	-72.7
KNN	33.45	3.63	0.56
RF	10.59	1.89	0.86
ANN	31.56	4.05	0.59
SVM	25.43	2.91	0.67
Average (without NB)	28.78	3.43	0.62

4.1.5 Disability Status

Table 11 displays the results of the evaluation metrics for the models trained to the disability status characteristics. The disability status models have similar results and pattern to the other characteristics. Although RF remains the highest performing algorithm this is its poorest performance. Whereas SVM performed its best with the disability status characteristic.

TABLE 11
EVALUATION METRICS RESULTS FOR MACHINE LEARNING ALGORITHMS FOR DISABILITY STATUS

Model	Evaluation Metrics		
	MSE	MAE	R^2
LR	45.04	4.62	0.48
NB	5725.73	75.19	-65.64
KNN	31.46	3.96	0.63
RF	13.61	2.09	0.84
ANN	25.37	3.55	0.7
SVM	15.1	2.06	0.82
Average (without NB)	26.12	3.26	0.69

4.1.6 Evaluation Metrics Summary

Table 12 describes the average performance of each machine learning algorithms. RF and SVM are the better performing models with RF being the highest performing model. ANN surprisingly performing averagely contrastingly to what some studies had suggested would happen [6], [7], [17].

TABLE 12
AVERAGE PERFORMANCE OF MACHINE LEARNING ALGORITHMS

Model	Average Evaluation Metrics Result		
	MSE	MAE	R^2
ANN	28.49	3.8	0.64
KNN	32.20	3.84	0.60
LR	43.18	4.65	0.46
NB	5721.60	75.20	-70.65
RF	11.65	1.98	0.86
SVM	20.74	2.62	0.73
Average (without NB)	27.25	3.38	0.66

A possible explanation for ANN's average performance may be that the high performance it had produced in previous research was majorly within classification tasks as this is the more popular technique used in predicting student performance or satisfaction [6]. Therefore, the results may suggest that ANN, for regression within the context of predicting student satisfaction, does not perform as highly as in classification tasks. Additionally, in a study evaluating LR, SVM, and RF within a regression task RF had been evaluated as high performing and the better algorithm [59]. Further reasoning RF's high performance within this study and over ANN.

Another possible reason is the feature selection method. The features were manually selected based on logical reasoning of their appropriateness to the task. Due to the limited number of features left after the manual feature selection no feature selection technique were run on the remainder of features and therefore their relevance was not evaluated. ANN has been studied to have one of the worse tolerances to irrelevant attributes and therefore the possible inclusion of irrelevant attributes can cause negative effect on the algorithm [60]. Contrastingly SVM and DT are tolerant to irrelevant attributes and still perform well. Therefore, the inclusion of some possible irrelevant attributes may reason ANN's poorer performance and RF's (an enhanced type of DT) and SVM's ability to still perform highly.

Overall table 12 contributes to answering RQ1 as RF is the highest performing algorithm when predicting student satisfaction based on student characteristics according to the evaluation metrics.

Tables 7-11 portray the completion of objective 1 and 3 of this study as it portrays the use of all five student characteristics to predict student satisfaction. As well all six ML algorithms described in objective 3 have been implemented and evaluated with the evaluation metrics.

4.1.8 Plots result

Append ix 4 includes all the graphs of the residual plots and prediction error plots for each model tested in detail. Figures 10 and 11 below give a visual overview of all the plots, residual and prediction error, aligned by characteristics and algorithms. This allows to visualise and analyse any patterns in performance.

From LRs prediction error plots implies is consistently predicting only high values of student satisfaction as the data points skews off the diagonal with lower satisfaction actual values. This helps understand LRs low performance metrics contrastingly to other model as it struggles with predicting lower satisfaction rates and lower extremes. The residual plots of LR are the most spread out too and are not at its highest density at the point of origin further implying the model not being a good fit for the task.

NB from its prediction error plots interestingly seems to be stuck with predicting target values at a low value and looking in appendix 4 it can be seen it only predicts values to be between 0 and 10. This may imply that because the data was discretised into ten-bins its currently only predicting data within the first bin further suggesting an issue with the discretisation method.

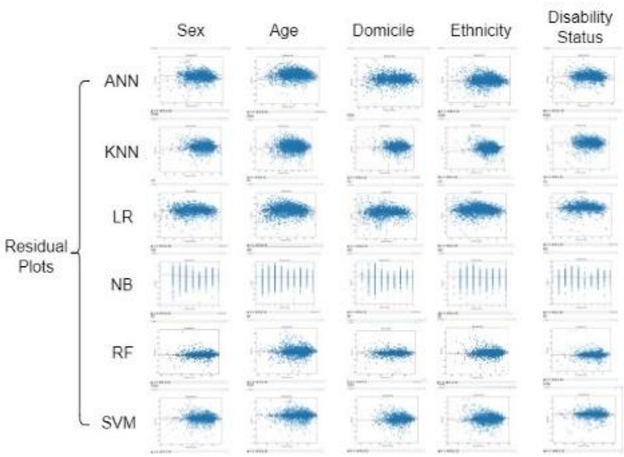


Fig 10. Visual display of all the residual plots for all characteristics and algorithms.

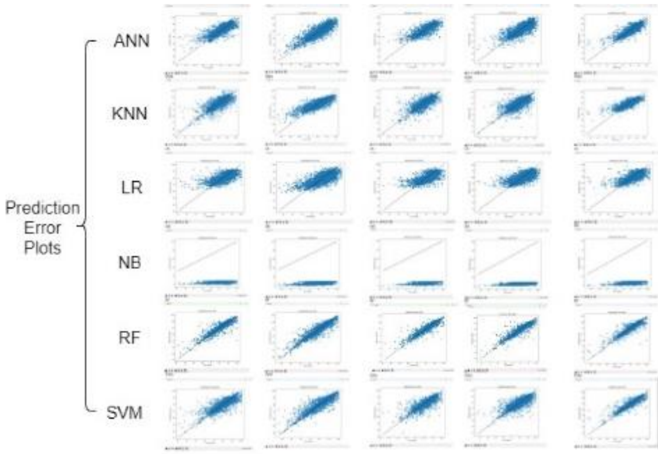


Fig 11. Visual display of all the prediction error plots for all characteristics and algorithms.

Both SVM and RF consistently start at the origin and produce almost full symmetry on the residual plots contrastingly to the other models once more implying their better performance and fit to the task. RF's prediction error plots across all characteristics have a very tight clustering and a narrow spread around the diagonal line contrastingly to other ML algorithm plots. This indicates that the predictions are consistently close to the actual values meaning the RF algorithm is highly accurate and has minimal error. Therefore, once more reemphasising that RF is the highest performing algorithm, answering RQ1.

4.2 Display Interface Results

Appendix 5 contains all the responses for the display interface feedback form filled by users during the beta testing. The feedback given from the interface is overall positive. For the interface itself all responses agreed to the ease of use, ease of navigation through the website and the website being user-friendly. As well, all users were at least somewhat satisfied by the features and functionality provided. The engagement levels varied but on average users were very or extremely engaged with the website. The information was averagely rated 4.86/5 as clearly displayed. The only issue that was reported was about trying to predict other characteristics and the predictions not changing but within the backend everything was reported as working properly which would mean the predictions really did not change.

Some improvements were suggested, one user suggest the layout of the interface could be made a bit nicer and more engaging. Two users mentioned that the predictions were not obviously laid out when displayed and that some did not notice they had been made and that they could be made to pop a bit more.

28.5% of users answered that yes, their predicted satisfactions were accurate, 43% of users answered that it was accurate in some themes and in others not, and 28.5% of users answered that their predicted satisfaction was not accurate. This suggests the model is not specifically as high performing in real life.

29% of users were not impacted at all by their prediction, 43% of users were not impacted by their prediction, 14% of users were neutral about if they were impacted or not, and 14% were impacted by the predictions. With 72% of users stated that they were not impacted by the predictions its mays suggest it is not relevant to them. However, some users that were not impacted had stated it was due to the warning message upon entering the website which suggests that the warning message was successful in its aim therefore if there had been no warning maybe it would have been relevant to them.

71% of users said yes, they would have benefited from a website as such to choose their degree. Multiple users mentioned that satisfaction would play a role in choosing their degree and one user citing its link to experience and the importance of these criteria's. 29% of users said they would have not benefited from this website to choose their degree as satisfaction was not the criteria, they used to pick their degree. This implies a need for a tool such as the developed prototype however it would not be applicable or utilised by everyone.

Overall, the questionnaire and feedback suggested that a display interface utilising the machine learning algorithms to predict student satisfaction based on characteristics is beneficial to the majority of real-life students thus answering RQ2.

5 Evaluation

5.1 Research Questions

RQ1 was: *What is the highest performing machine learning algorithm, according to evaluation metrics and diagnostic plots, to predict student satisfaction based on student characteristics?*

Within the research and study conducted the results of the evaluation metrics, as seen in table 12, and diagnostic plots, as seen in figure 10 and 11, identified RF as the highest performing algorithm to predict student satisfaction based on characteristics. This was contradictory to a majority of the literature analysed apart from one research which had stated RF as the highest performing algorithm [9]. As although RF had been suggested as a good performing algorithm, ANN and SVM had been identified as the highest performing algorithms within literature. SVM was still observed to have the second highest performance. ANN performed averagely which was surprising based on various literature suggesting it would be the highest performing algorithm [6], [7], [17]. ANN's low performance may be explained by the limited feature selection method that may have caused the involvement of irrelevant feature which ANN has poor tolerance of and therefore its performance is negatively impacted [60].

Additionally, RF's high performance did not overly improve after hyper tuning as seen in table 13, suggesting that the initial configuration of the model's hyperparameters is already well-suited to the task of predicting student satisfaction based on student characteristics.

This result brings in a new perspective within the field and potentially suggests that RF is a high performing model to predict student satisfaction based on characteristics as a regression-based task.

RQ2 was: *How beneficial and relevant is a display interface utilising the machine learning algorithms to predict student satisfaction based on characteristic to real life students?*

The results of the feedback form implied that the display interface overall would be beneficial to students and potential students. Although the limited number of responses require more research around this question. To determine the relevance of the display interface further research without the warning sign is required and further questioning on the relevance of such an interface is also required.

5.2 Objectives

As a reminder the objectives of this study were:

Objective 1: Use as predictive features at least three of the following characteristics to help predict student satisfaction: sex, age, domicile, ethnicity, and disability.

Objective 2: Use at least four different satisfaction themes from the seven NSS themes.

Objective 3: Implement and evaluate at least three of the following machine learning algorithms: Artificial Neural Network, K-Nearest Neighbour, Linear Regression, Naïve Bayes, Random Forest, and Support Vector Machine.

Objective 4: Create and implement an interactive display interface that will allow student to input their characteristics and it will output the student's satisfaction.

All objectives were fulfilled at their fullest with their methodologies and results described previously.

5.3 Machine Learning Models

The models amid NB all worked on each characteristic in predicting student satisfaction with RF being identified as the better performing machine learning model. SVM also being a suitably performing algorithm for predicting student satisfaction based on characteristics.

The methodology used to construct the models was suitable and well built. The preprocessing helped limit any issues such as overfitting and the hyper tuning methods helped optimise the models. Table 13 helps further evaluate the hyperparameters tuning process. Table 13 depicts the average MSE evaluation metric of each model on validation sets pre and post hyper tuning. The full pre and post MSE result per characteristic per model can be observed in Appendix 7.

TABLE 13
PRE VS POST HYPER TUNING VS TEST MSE OF MACHINE LEARNING ALGORITHMS

	Pre	Post	Test	Difference pre vs post
ANN	43.88	30.59	28.49	13.29
KNN	38.08	33.6	32.20	4.48
LR	45.31	45.38	43.18	-0.07
NB	8.73	7.02	5721.60	1.71
RF	13.19	13.07	11.65	0.12
SVM	36.11	21.92	20.74	14.19
Average	30.88	25.26	27.25	5.62

The hyper tuning process can be seen to have been beneficial and effective to most models with both ANN and SVM benefitting the most. Interestingly RF had very little difference made by the hyper tuning suggesting that the model was already performing highly with the default hyperparameters and therefore a good fit for the task.

The hyperparameter tuning for LR actually made the model perform slightly worse on average on the validation sets. This suggests that the tuning method which involved elastic net regularisation suggested by [55] might not have been effective in optimizing the LR performance. However, looking at the test set data the LR performed better on average than its validation sets suggesting that the tuning process did possibly have a positive impact on LR’s generalisation performance. The fact that LR's performance on the test data was superior post-tuning indicates that the selected hyperparameters likely resulted in a model that generalized well to unseen data.

The model’s performance on the validation sets were only evaluated with MSE and having had issues such as the overfitting of NB it highlights the importance of not relying solely on one metric during the training phase. As had R-squared been used to evaluate NB during training there is a chance the discretisation error would have been caught and fixed before the model was exposed to the test set.

Another potential improvement is with the method of application to weight the response number and manage imbalances in the data. Class weights was used which seemed to have worked however it would have been more suitable to try multiple technique and find the better performing technique. Additionally SMOTE could have been used as a more traditional method to further manage imbalances within the data in [44].

The method used to discretise NB led to severe overfitting and was not suitable. It would be interesting for future research to attempt other discretisation techniques on the model to find a suitable one or potentially try an entropy-based method as it was suggested to perform slightly better than other discretisation methods [52].

The model performance was solely evaluated and determined based on the evaluation metrics used (MSE, MAE, R-squared) and the plots (residual plot and prediction error plot). The ML algorithms performance could further be analysed and determined based on running efficiency, computational efficiency, and other potential determiners of ML performance. Therefore, further research could be conducted to evaluate the highest performing ML algorithm based on different performance factors.

5.3 Display Interface

The interface prototype is working well although very simplistic. Additionally, the general response to such an interface from users was overall positive many indicating the usefulness and benefice such an interface would bring. However, the interface could have been better designed to be more engaging as suggested from the feedback.

The interface was only a prototype and is limited to running locally meaning that for the beta testing convenience sampling, based on people able to access the local device, had to be used. This limited the number of people to try it and give feedback. It also limited the diversity of the users within the convenience sample as they were all from the same university causing a potential bias in some results within the feedback form. Only seven users took part in the beta testing which limits the validity of the results on the feedback form as the response numbers are low. In turn this questions the validity of the answer to RQ2. It would therefore be beneficial to increase the sample size to pick from to have more users take part in the beta testing.

For the interface to be functional more than just locally it would need to have been adapted within its backend so it could be an up and running website. Additionally due to the nature of needing to train a model for each characteristic during the launch of the website, the website takes a long time to load. Time efficiencies of models was not considered during evaluation even though it is crucial to the solution as the model needs the ability to train relatively quick in order not to cause to many delays for the user using the interface.

The ethical implication of such an interface also should be evaluated. Even though the students are warned currently that the student satisfaction predictions are not realistic there are many implications of developing a potential interface as such that would be realistic. The use of machine learning to predict student satisfaction raises concerns about not only the accuracy but also the fairness of the predictions as well as the potential for biased decision making [27]. Furthermore, there is a risk that machine learning algorithms may not fully understand the complexities of the educational context, leading to inaccurate predictions and potentially harmful decisions [27]. The use of machine learning in predicting student satisfaction may raise questions about algorithmic accountability. It may be challenging to hold the algorithms accountable for their predictions, especially in cases where the educational domain is not sufficiently understood [27] or where there is a lack of consensus on the metrics used to measure student satisfaction. This lack of accountability can lead to decisions being made based on opaque and potentially flawed predictions, which could have negative consequences for students.

5.4 Data Limitations

A limitation to the solution and the study is the data used. Although the data is suitable it was not the ideal type of data for the study. Ideally the data would be a dataset of all the individual NSS 2023 responses with each students’ characteristics. This would have created a more suitable solution for the task as it would be creating models based on all demographic characteristics and university related characteristics at once rather than the characteristics individually. This would have also helped models be more accurate in their prediction to users as the increased features would allow for more tailored and accurate predictions within the interface [39]. More meaningful patterns between characteristics and machine learning may have been learned by the model and potentially the results of the better performing machine learning model may have differed.

Additionally, the limited data exploration conducted limited the understanding of the data and its trends which could have aided in better tailoring the models to the data. Literature suggests that the NSS data needs to be well understood and examined before using it for predictions as research suggest a risk of using machine learning in education without a deeper understanding of the data [27].

5.5 Project Organisation

The strength within the project management was the ability to adapt and set weekly goals to allow incremental progress every week. Although slow at first once the study found its direction progress was made at a suitable pace. The studies initial aim was around predicting student experience however due to the limited data within this field the studies focus was shifted to student satisfaction, a subset of student experience.

In the occasional cases that progress was slowed, and delays were caused on the plan this was mostly due to external factors in which case the plan had been adapted adequately to fit new timeframes.

The study could have benefited from better planned research around the topic and study for a more efficient execution of the methodology.

6 Conclusion

6.1 Main Finding

The highest performing machine learning algorithm to predict student satisfaction based on student characteristics (sex, age, domicile, ethnicity, and disability) was RF. Notably SVM also demonstrated strong performance particularly in certain characteristics like disability status and age. The high performance was based on the models results on the evaluation metrics (MSE, MAE, R-squared) and the plots (residual plot and prediction error plot). Surprisingly ANN performed averagely contrastingly to previous studies suggestion of its high performance within the field [6], [7], [17]. This was suggested due to the potential limitation within the feature selection method leading to possible involvement of irrelevant attributes to which the algorithm is sensitive to [60]. The hyperparameter tuning process, while beneficial for most models, showed marginal improvement for RF, indicating that its default configuration is well-suited for the task.

From the model's higher average performance with the student characteristic age, it became evident that age emerged as a more influential demographic characteristic in predicting student satisfaction compared to sex, contradicting previous literature. This shift in perspective highlights the importance of considering various demographic factors and their unique impacts on satisfaction levels.

An implementation of the model into a website display interface so that students may interact with it was found beneficial. This contributes new research to the field of predicting student satisfaction as it discusses and evaluates the use of different machine learning algorithm, it builds and predicts the algorithm using NSS 2023 data which had not been researched before.

While the project achieved its goals, there were notable limitations, such as the data's nature and exploration, which may have affected the models' accuracy and generalizability. Additionally, ethical considerations surrounding algorithmic accountability and fairness in predictive models underscore the need for cautious implementation and ongoing evaluation.

In conclusion, this project contributes novel evidence supporting Random Forest (RF) as a high-performing algorithm for predicting student satisfaction based on student characteristics. Additionally, the development and evaluation of a display interface offer promising insights into the potential benefits of allowing students to predict their satisfaction levels. While these findings are significant, further research and validation are warranted to fully assess the efficacy and broader implications of such predictive models and interfaces in educational contexts.

6.2 Further Works

In future research, the machine learning methodology should be extended by leveraging the more comprehensive NSS 2023 datasets containing individual responses. This extension could integrate additional factors identified as influential in student satisfaction, such as demographic, academic, and environmental variables. Incorporating these factors into the model could yield a more realistic representation of student satisfaction, potentially enhancing both its applicability and accuracy within the UK higher education context. As well as further exploration of different hyper tuning methods or the evaluation of different performance factors.

Regarding the interface, future efforts could focus on refining its design based on user feedback and usability principles. Specific enhancements may include improving layout and visual appeal, as well as incorporating additional features to enhance user engagement and interaction. Furthermore, investigation of the broader implications of such a predictive model, considering ethical considerations such as transparency, fairness, and algorithmic accountability, would be advisable.

While these advancements hold promise for enhancing decision-making in higher education, it's important to recognize the potential challenges that may arise in implementation, such as data availability, computational resources, and ethical concerns. Proactively addressing these challenges and refining methodologies could lead to meaningful contributions to the field of predictive modelling in educational contexts.

REFERENCES

[1] P. Bolton, "Higher Education Student Numbers," House of Commons Library, Jan. 2024. Available: https://researchbriefings.files.parliament.uk/documents/CBP-7857/CBP-7857.pdf

[2] S. Gibbons, E. Neumayer, and R. Perkins, "Student satisfaction, league tables and university applications: Evidence from Britain," Economics of Education Review, vol. 48, pp. 148–164, Oct. 2015, doi: https://doi.org/10.1016/j.econedurev.2015.07.002. Available: https://www.sciencedirect.com/science/article/pii/S0272775715000862

[3] Office for Students, "National Student Survey - NSS - Office for Students," Officeforstudents.org.uk, 2020. Available: https://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/

[4] M. D. Clemes, C. E. C. Gan, and T.-H. Kao, "University Student Satisfaction: An Empirical Analysis," Journal of Marketing for Higher Education, vol. 17, no. 2, pp. 292–325, Apr. 2008, doi: https://doi.org/10.1080/08841240801912831. Available: https://www.tandfonline.com/doi/full/10.1080/08841240801912831

[5] M. Pojon, "Using Machine Learning to Predict Student Performance," University of Tampere , 2017. Available: https://urn.fi/URN:NBN:fi:uta-201706262111

[6] Y. Baashar, G. Alkawsi, N. Ali, H. Alhussian, and H. T. Bahbouh, "Predicting student's performance using machine learning methods: A systematic literature review," IEEE Xplore, Jul. 01, 2021. doi: https://doi.org/10.1109/ICCOINS49721.2021.9497185. Available: https://ieeexplore.ieee.org/abstract/document/9497185.

[7] H. Altabrawee, O. A. J. Ali, and S. Q. Ajmi, "Predicting Students' Performance Using Machine Learning Techniques," JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences, vol. 27, no. 1, pp. 194–205, Apr. 2019, doi: https://doi.org/10.29196/jubpas.v27i1.2108. Available: https://www.researchgate.net/publication/332893829_Predicting_Students'_Performance_Using_Machine_Learning_Techniques

[8] L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review," Applied Sciences, vol. 10, no. 3, p. 1042, Feb. 2020, doi: https://doi.org/10.3390/app10031042. Available: https://www.mdpi.com/2076-3417/10/3/1042

[9] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," Computers & Education, vol. 131, pp. 22–32, Apr. 2019, doi: https://doi.org/10.1016/j.compedu.2018.12.006. Available: https://www.sciencedirect.com/science/article/pii/S0360131518303191?casa_token=kw7P_-CScV0AAAAA:OlShs6yiOYcyAJQQnt_PRPfcAlXHU2MEHhnWJOlxWef0vcsA_fC5dfzIn2hEzYjm1p-Y0b8IXw.

[10] J. Fernandez-Garcia, J. C. Preciado, F. Melchor, R. Rodriguez-Echeverria, J. M. Conejero, and F. Sanchez-Figueroa, "A Real-Life Machine Learning Experience for Predicting University Dropout at Different Stages Using Academic Data," IEEE Access, vol. 9, pp. 133076–133090, 2021, doi: https://doi.org/10.1109/access.2021.3115851. Available: https://ieeexplore.ieee.org/abstract/document/9548895

[11] Mehrdad Arashpour et al., "Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization," Computer Applications in Engineering Education, vol. 31, no. 1, pp. 83–99, Sep. 2022, doi: https://doi.org/10.1002/cae.22572. Available: https://doi.org/10.1002/cae.22572

[12] Kanwar and M. Sanjeeva, "Student satisfaction survey: a key for quality improvement in the higher education institution," Journal of Innovation and Entrepreneurship, vol. 11, no. 1, Mar. 2022, doi: https://doi.org/10.1186/s13731-022-00196-6. Available: https://innovation-entrepreneurship.springeropen.com/articles/10.1186/s13731-022-00196-6

[13] M. Nurunnabi and A. Abdelhadi, "Student satisfaction in the Russell Group and Non-Russell Group Universities in UK," Data in Brief, vol. 22, pp. 76–82, Feb. 2019, doi: https://doi.org/10.1016/j.dib.2018.11.124. Available: https://reader.elsevier.com/reader/sd/pii/S2352340918315178?token=F34727D7741E9A8D78445FDF278FADAF99568A553FA50F5C178AD257FDB7DF23513DB66CC8FE4AEB58337B40A146BEB5

[14] V. Skrbinjek and V. Dermol, "Predicting students' satisfaction using a decision tree," Tertiary Education and Management, vol. 25, no. 2, pp. 101–113, Jan. 2019, doi: https://doi.org/10.1007/s11233-018-09018-5. Available: https://link.springer.com/article/10.1007%2Fs11233-018-09018-5

[15] W. H. Wong and E. Chapman, "Student satisfaction and interaction in higher education," Higher Education, vol. 85, Jun. 2022, doi: https://doi.org/10.1007/s10734-022-00874-0. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9159046/

[16] M. K. Ho, K. Y. Cheong, and A. Weldon, "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques," PLOS ONE, vol. 16, no. 4, p. e0249423, Apr. 2021, doi: https://doi.org/10.1371/journal.pone.0249423. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249423

[17] D. Kamal, "Using Artificial Neural Network to Predicted Student Satisfaction in E-learning," American Journal of Applied Mathematics and Statistics, vol. 8, no. 3, pp. 90–95, Sep. 2020, doi: https://doi.org/10.12691/ajams-8-3-2. Available: http://article.sciappliedmathematics.com/pdf/AJAMS-8-3-2.pdf

[18] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach," Computers & Education, vol. 145, p. 103724,

Feb. 2020, doi: https://doi.org/10.1016/j.compedu.2019.103724. Available: https://www.sciencedirect.com/science/article/pii/S0360131519302775

[19] Tovmasyan, D. Walker, and L. Kaye, "Can Personality Traits Predict Students' Satisfaction with Blended Learning during the COVID-19 Pandemic?," College Teaching, pp. 1–7, Dec. 2022, doi: https://doi.org/10.1080/87567555.2022.2156450. Available: https://www.tandfonline.com/doi/full/10.1080/87567555.2022.2156450

[20] K. M. Elliott and M. A. Healy, "Key Factors Influencing Student Satisfaction Related to Recruitment and Retention," Journal of Marketing for Higher Education, vol. 10, no. 4, pp. 1–11, Jun. 2001, doi: https://doi.org/10.1300/j050v10n04_01.Available: https://www.tandfonline.com/doi/abs/10.1300/J050v10n04_01

[21] F.-S. Wach, J. Karbach, S. Ruffing, R. Brünken, and F. M. Spinath, "University Students' Satisfaction with their Academic Studies: Personality and Motivation Matter," Frontiers in Psychology, vol. 7, Feb. 2016, doi: https://doi.org/10.3389/fpsyg.2016.00055. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4754397/

[22] Department for Business Innovation & Skills, "The Benefits of Higher Education Participation for Individuals and Society: key findings and reports 'The Quadrants' ," Oct. 2013. Available: https://assets.publishing.service.gov.uk/media/5a7ba5fee5274a7318b9004a/bis-13-1268-benefits-of-higher-education-participation-the-quadrants.pdf

[23] H. J. Green, M. Hood, and D. L. Neumann, "Predictors of Student Satisfaction with University Psychology Courses: A Review," Psychology Learning & Teaching, vol. 14, no. 2, pp. 131–146, Jul. 2015, doi: https://doi.org/10.1177/1475725715590959. Available: https://journals.sagepub.com/doi/epub/10.1177/1475725715590959

[24] Denson, T. Loveday, and H. Dalton, "Student evaluation of courses: what predicts satisfaction?," Higher Education Research & Development, vol. 29, no. 4, pp. 339–356, Aug. 2010, doi: https://doi.org/10.1080/07294360903394466. Available: https://www.tandfonline.com/doi/full/10.1080/07294360903394466

[25] K. M. Elliott and D. Shin, "Student Satisfaction: An alternative approach to assessing this important concept," Journal of Higher Education Policy and Management, vol. 24, no. 2, pp. 197–209, Nov. 2002, doi: https://doi.org/10.1080/1360080022000013518. Available: https://www.tandfonline.com/doi/abs/10.1080/1360080022000013518

[26] L. Sweeney, "A Predictive Model of Student Satisfaction," DOAJ (DOAJ: Directory of Open Access Journals), vol. 5, no. 1, Jan. 2016, doi: https://doi.org/10.21427/d7mh80. Available: https://arrow.tudublin.ie/ijap/vol5/iss1/8/

[27] D. Parapadakis, "Can artificial intelligence help predict a learner's needs? Lessons from predicting student satisfaction," London Review of Education, vol. 18, no. 2, Jul. 2020, doi: https://doi.org/10.14324/lre.18.2.03. Available: https://eric.ed.gov/?id=EJ1297539

[28] Kwok Tong Soo, "Student satisfaction and degree performance," Jan. 2010, Available: https://www.lancaster.ac.uk/people/sook/papers/PFunction9.pdf

[29] H. Gull, M. Saqib, S. Z. Iqbal, and S. Saeed, "Improving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning," 2020 IEEE International Conference for Innovation in Technology (INOCON), Nov. 2020, doi: https://doi.org/10.1109/inocon50539.2020.9298266. Available: https://ieeexplore.ieee.org/abstract/document/9298266?casa_token=7WlzESHDE_4AAAAA:qE4XqzJv3f19YV4AfILHYwg9eqZ_Mr8-wtIU3rQpgpfmY_4UjBPjeMzX-G9NcH5hHQhqK4a2fLo

[30] N. Matus, C. Rusu, and S. Cano, "Student eXperience: A Systematic Literature Review," Applied Sciences, vol. 11, no. 20, p. 9543, Oct. 2021, doi: https://doi.org/10.3390/app11209543. Available: https://www.mdpi.com/2076-3417/11/20/9543

[31] C. McINNIS, "Studies of Student Life: an overview," European Journal of Education, vol. 39, no. 4, pp. 383–394, Dec. 2004, doi: https://doi.org/10.1111/j.1465-3435.2004.00192.x. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1465-3435.2004.00192.x?saml_referrer

[32] R. Maity, P. P. Bhagwat, and A. Bhatnagar, "Potential of support vector regression for prediction of monthly streamflow using endogenous property," Hydrological Processes, vol. 24, no. 7, pp. 917–923, Mar. 2010, doi: https://doi.org/10.1002/hyp.7535. Available: https://onlinelibrary.wiley.com/doi/epdf/10.1002/hyp.7535?saml_referrer

[33] E. Staddon and P. Standish, "Improving the Student Experience," Education Policy: Philosophical Critique, pp. 118–128, Apr. 2013, doi: https://doi.org/10.1002/9781118680544.ch10. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118680544.ch10

[34] S. Rong and Z. Bao-wen, "The research of regression model in machine learning field," MATEC Web of Conferences, vol. 176, p. 01033, 2018, doi: https://doi.org/10.1051/matecconf/201817601033. Available: https://www.matec-conferences.org/articles/matecconf/pdf/2018/35/matecconf_ifid2018_01033.pdf

[35] A. Alwosheel, S. van Cranenburgh, and C. G. Chorus, "Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis," Journal of Choice Modelling, vol. 28, pp. 167–182, Sep. 2018, doi: https://doi.org/10.1016/j.jocm.2018.07.002. Available: https://www.sciencedirect.com/science/article/pii/S1755534518300058?casa_token=OtTITSJS4cIAAAAA:bwd4fU

PlcXL3YFucBo77wIUtfnjuUVF50LuwPGZ3sboo8ZCptAs-AUU_yvzxFe5nLPme3L8

[36] S. Nallathambi and K. Ramasamy, "Prediction of electricity consumption based on DT and RF: An application on USA country power consumption," in IEEE Xplore, Apr. 2017, pp. 1–7. doi: https://doi.org/10.1109/ICEICE.2017.8191939. Available: https://ieeexplore.ieee.org/document/8191939

[37] Office for Students, "Download the NSS data - Office for Students," www.officeforstudents.org.uk, Aug. 10, 2023. Available: https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/download-the-nss-data/

[38] Office for Students, "Home - Office for Students," Officeforstudents.org.uk, 2019. Available: https://www.officeforstudents.org.uk/

[39] L. Jollans et al., "Quantifying performance of machine learning methods for neuroimaging data," NeuroImage, vol. 199, pp. 351–365, Oct. 2019, doi: https://doi.org/10.1016/j.neuroimage.2019.05.082. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6688909/

[40] Pandas, "Python Data Analysis Library," Pydata.org, 2018. Available: https://pandas.pydata.org/

[41] "scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation," scikit-learn.org. Available: https://scikit-learn.org/stable/index.html#

[42] S. Raschka, J. Patterson, and C. Nolet, "Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence," Information, vol. 11, no. 4, p. 193, Apr. 2020, doi: https://doi.org/doi.org/10.3390/info11040193. Available: https://www.mdpi.com/2078-2489/11/4/193

[43] J. Hao and T. K. Ho, "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language," Journal of Educational and Behavioral Statistics, vol. 44, no. 3, p. 107699861983224, Feb. 2019, doi: https://doi.org/10.3102/1076998619832248. Available: https://journals.sagepub.com/doi/epub/10.3102/1076998619832248

[44] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," Journal of Reliable Intelligent Environments, vol. 7, pp. 263–275, Jan. 2021, doi: https://doi.org/10.1007/s40860-021-00133-6. Available: https://link.springer.com/article/10.1007/s40860-021-00133-6

[45] T. Kessler, G. Dorian, and J. H. Mack, "Application of a Rectified Linear Unit (ReLU) Based Artificial Neural Network to Cetane Number Predictions," Volume 1: Large Bore Engines; Fuels; Advanced Combustion, Oct. 2017, doi: https://doi.org/10.1115/icef2017-3614. Available: https://asmedigitalcollection.asme.org/ICEF/proceedings-abstract/ICEF2017/58318/245474

[46] R. K. Yadav and Anubhav, "PSO-GA based hybrid with Adam Optimization for ANN training with application in Medical Diagnosis," Cognitive Systems Research, vol. 64, pp. 191–199, Dec. 2020, doi: https://doi.org/10.1016/j.cogsys.2020.08.011. Available: https://www.sciencedirect.com/science/article/pii/S1389041720300589?casa_token=7YU06DFoK-cAA-AAA:maxsCrepATpWnnqnaifhAleApwPG5i8KwQLO-_9M_zdNhJHfsnEoXGHqtFpSbnL5I4sLuJ4

[47] M. Steurer, R. J. Hill, and N. Pfeifer, "Metrics for evaluating the performance of machine learning based automated valuation models," Journal of Property Research, vol. 38, no. 2, pp. 99–129, Apr. 2021, doi: https://doi.org/10.1080/09599916.2020.1858937. Available: https://www.tandfonline.com/doi/full/10.1080/09599916.2020.1858937?scroll=top&needAccess=true

[48] W. Pannakkong, K. Thiwa-Anont, K. Singthong, P. Parthanadee, and J. Buddhakulsomsiri, "Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN," Mathematical Problems in Engineering, vol. 2022, pp. 1–17, Jan. 2022, doi: https://doi.org/10.1155/2022/8513719. Available: https://www.hindawi.com/journals/mpe/2022/8513719/

[49] D. M. Belete and M. D. Huchaiah, "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results," International Journal of Computers and Applications, vol. 44, no. 9, pp. 1–12, Sep. 2021, doi: https://doi.org/10.1080/1206212x.2021.1974663. Available: https://www.tandfonline.com/doi/abs/10.1080/1206212X.2021.1974663

[50] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," Neurocomputing, vol. 415, pp. 295–316, Nov. 2020, doi: https://doi.org/10.1016/j.neucom.2020.07.061. Available: https://www.sciencedirect.com/science/article/pii/S0925231220311693

[51] R. Hossain and D. Timmer, "Machine Learning Model Optimization with Hyper Parameter Tuning Approach," Global journal of computer science and technology, vol. 21, no. 2, Sep. 2021, Available: https://gjcst.com/index.php/gjcst/article/view/255

[52] C.-N. Hsu, H.-J. Huang, and T.-T. Wong, "Implications of the Dirichlet Assumption for Discretization of Continuous Variables in Naive Bayesian Classifiers," Machine Learning, vol. 53, no. 3, pp. 235–263, 2003, doi: https://doi.org/10.1023/a:1026367023636. Available: https://link.springer.com/article/10.1023%2FA%3A1026367023636.

[53] R. Gupta, S. D. Sharma, and Y. Kumar, "Heart Disease Prediction using Machine learning algorithms," Jaypee University of Information Technology, Solan, H.P., 2023. Available: http://www.ir.juit.ac.in:8080/jspui/handle/123456789/9936

[54] H. Kaneko and K. Funatsu, "Fast optimization of hyperparameters for support vector regression models with highly predictive ability," Chemometrics and Intelligent Laboratory Systems, vol. 142, pp. 64–69, Mar. 2015, doi: https://doi.org/10.1016/j.chemolab.2015.01.001. Availa-

ble: https://www.sciencedirect.com/science/article/pii/S0169743915000039

[55] L. P. Coelho and W. Richert, Building machine learning systems with Python : get more from your data through creating practical machine learning systems with Python, 1st ed. Birmingham, UK : Packt Pub., 2015, p. 195. Available: https://ebookcentral.proquest.com/lib/durham/detail.action?docID=1236045&pq-origsite=primo

[56] M. A. Al-Alwani, L. Britt, S. Dunn-Norman, H. H. Alkinani, A. T. Al-Hameedi, and A. Al-Attar, "Production Performance Estimation from Stimulation and Completion Parameters Using Machine Learning Approach in the Marcellus Shale," onepetro.org, Jun. 2019, Available: https://onepetro.org/ARMAUSRMS/proceedings/ARMA19/All-ARMA19/ARMA-2019-2034/125065

[57] H. Yang, "Visual Assessment of Residual Plots in Multiple Linear Regression: A Model-Based Simulation Perspective ," Multiple Linear Regression Viewpoints, vol. 38, no. 2, Jan. 2012, Available: https://www.researchgate.net/publication/279751065_Visual_Assessment_of_Residual_Plots_in_Multiple_Linear_Regression_A_Model-Based_Simulation_Perspective

[58] U. B. Mogensen, H. Ishwaran, and T. A. Gerds, "Evaluating Random Forests for Survival Analysis Using Prediction Error Curves," Journal of Statistical Software, vol. 50, no. 11, 2012, doi: https://doi.org/10.18637/jss.v050.i11. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194196/

[59] Diyana Kinaneva, G. Hristov, Petko Kyuchukov, G. Georgiev, Plamen Zahariev, and Rosen Daskalov, "Machine Learning Algorithms for Regression Analysis and Predictions of Numerical Data," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Jun. 2021, doi: https://doi.org/10.1109/hora52670.2021.9461298. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9461298

[60] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," International Journal of Computer Trends and Technology, vol. 48, no. 3, pp. 128–138, Jun. 2017, doi: https://doi.org/10.14445/22312803/ijctt-v48p126. Available: https://www.researchgate.net/profile/J-E-T-Akinsola/publication/318338750_Supervised_Machine_Learning_Algorithms_Classification_and_Comparison/links/596481dd0f7e9b819497e265/Supervised-Machine-Learning-Algorithms-Classification-and-Comparison.pdf

APPENDICES

Appendix 1

1. Age
2. Care experienced students (Scotland)
3. Care experienced students (England, Wales, Northern Ireland)
4. Disability status
5. Disability type
6. Domicile
7. Estrangement
8. Ethnicity
9. Eligibility for free school meals
10. English Index of Multiple Deprivation 2019
11. Northern Ireland Index of Multiple Deprivation Measure 2017
12. Scottish Index of Multiple Deprivation 2020
13. Welsh Index of Multiple Deprivation 2019
14. Local students
15. Parental higher education
16. Service child status (at key stage 4)
17. Sex
18. Sexual orientation
19. TUNDRA (Tracking underrepresentation by area)

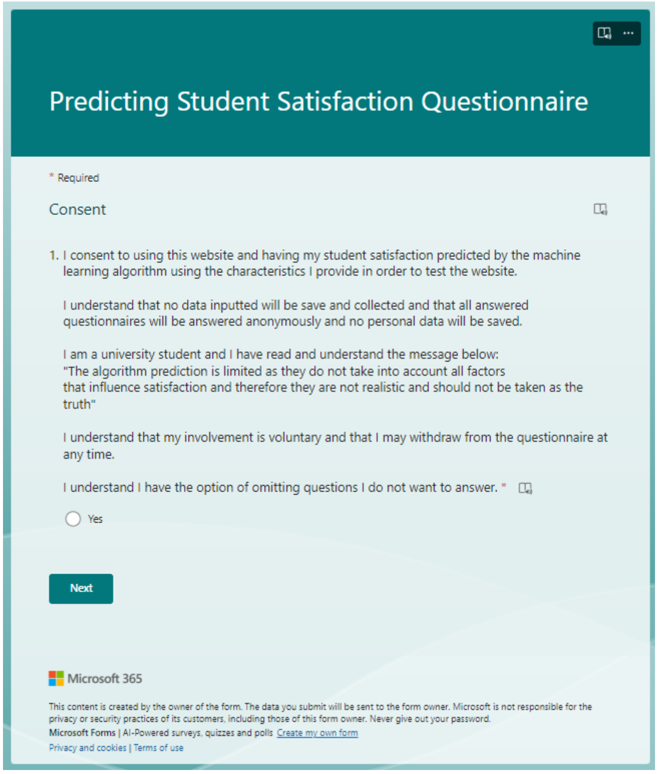
Appendix 2

Split: The values for the given characteristic
Provider country: England, Northern Ireland, Scotland, Wales, UK
Mode of study: Full-time, Part-time, Apprenticeship, All modes
Level of study: First degree, Other undergraduate, Undergraduate with postgraduate component, All undergraduates
Subject code: CAH subject code
Subject: CAH subject name
Question: NSS 2023 Question/Theme text
Responses: Full person equivalence number of responses to each question (excluding those who responded 'This does not apply to me').
Population: Full person equivalence number of eligible students
Suppression reason: DPL shown in the suppression row to represent "data protection, low". DP shown in the suppression row represents a secondary suppression for theme measures where one of the questions has been DPL suppressed. BK means the benchmark has been suppressed because there was a benchmarking factor with too many unknowns.
Option 1-5: Represents the number of respondents (weighted by full person equivalence) who selected each response option. Option 1 is the most positive option for a question, Option 4 is the most negative option for the four-point response scale. Option 5 is the most negative response option for question 28.
N/A: Number of respondents who selected 'This does not apply to me' (weighted by full person equivalence)
Positivity measure: Positivity Measure (percentage), pro-

portion of responses that were positive
Benchmark: Benchmarks reflect the sector average positivity measure but adjusted to reflect the mix of students and subjects at the provider. The adjustment takes account of the following factors (subject of study, level of study, age, sex, ethnicity, disability, and the mode of study).
Difference: Difference between positivity measure and benchmark
Standard deviation: Standard deviation of the difference between the positivity measure and the benchmark
Contribution to benchmark: Percentage that provider contributes to their own benchmark. A high contribution to benchmark makes it more likely that results will be close to the benchmark and makes the result less meaningful
Materially below benchmark: Proportion of the difference from benchmark statistical uncertainty distribution which is below -2.5
In line with benchmark: Proportion of the difference from benchmark statistical uncertainty distribution which is between -2.5 and 2.5
Materially above benchmark: Proportion of the difference from benchmark statistical uncertainty distribution which is above 2.5
Publication response headcount: Publication respondent headcount for survey as a whole (including those who responded 'This does not apply to me'). This field is used to apply the publication threshold of ten respondents.
Publication response rate: Publication response rate for survey as a whole (including those who responded 'This does not apply to me'). This field is used to apply the publication threshold of 50% response rate.

Appendix 3

Question 1:



Predicting Student Satisfaction Questionnaire

* Required

Consent

1. I consent to using this website and having my student satisfaction predicted by the machine learning algorithm using the characteristics I provide in order to test the website.

I understand that no data inputted will be save and collected and that all answered questionnaires will be answered anonymously and no personal data will be saved.

I am a university student and I have read and understand the message below:
"The algorithm prediction is limited as they do not take into account all factors that influence satisfaction and therefore they are not realistic and should not be taken as the truth"

I understand that my involvement is voluntary and that I may withdraw from the questionnaire at any time.

I understand I have the option of omitting questions I do not want to answer. *

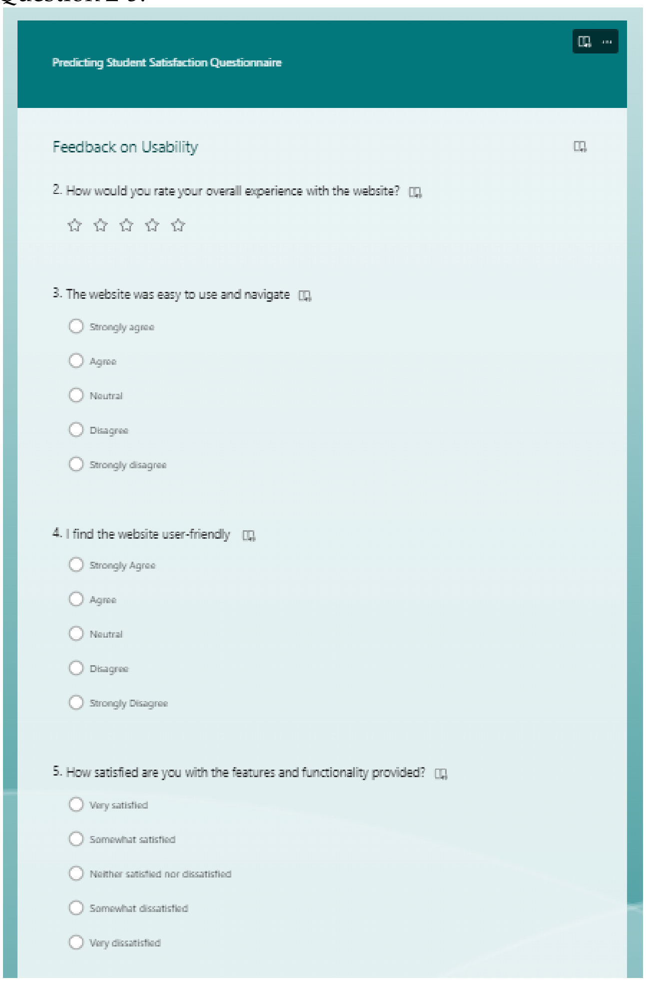
Yes

Next

Microsoft 365

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.
Microsoft Forms | AI-Powered surveys, quizzes and polls Create my own form
Privacy and cookies | Terms of use

Question 2-5:



Predicting Student Satisfaction Questionnaire

Feedback on Usability

2. How would you rate your overall experience with the website?

3. The website was easy to use and navigate

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

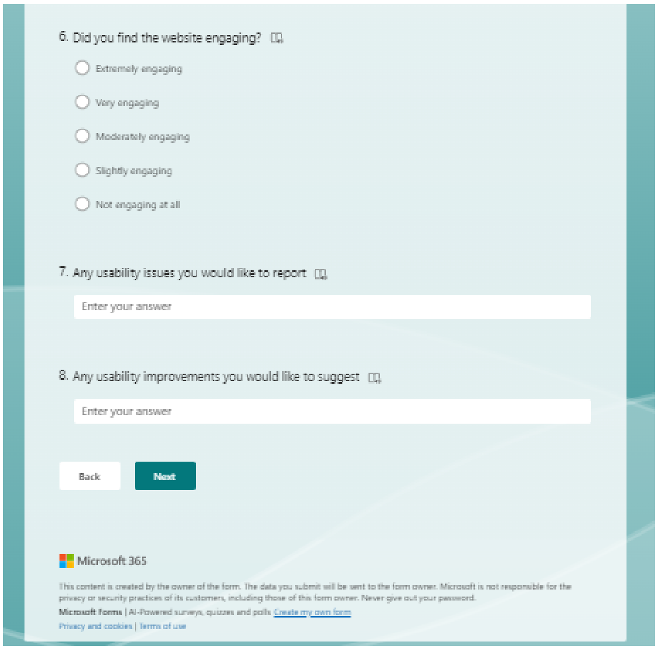
4. I find the website user-friendly

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

5. How satisfied are you with the features and functionality provided?

- Very satisfied
- Somewhat satisfied
- Neither satisfied nor dissatisfied
- Somewhat dissatisfied
- Very dissatisfied

Question 6-8:



6. Did you find the website engaging?

- Extremely engaging
- Very engaging
- Moderately engaging
- Slightly engaging
- Not engaging at all

7. Any usability issues you would like to report

Enter your answer

8. Any usability improvements you would like to suggest

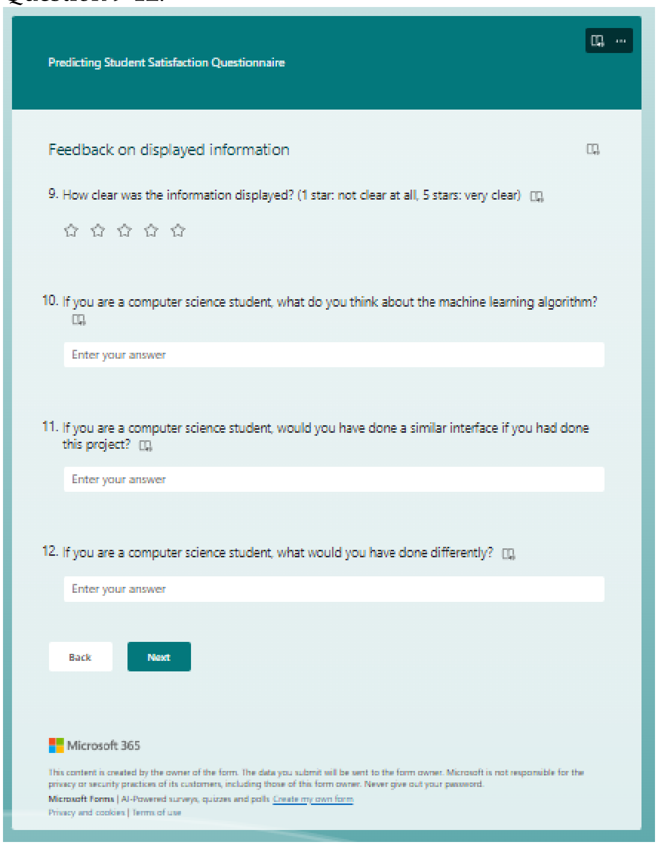
Enter your answer

Back Next

Microsoft 365

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.
Microsoft Forms | AI-Powered surveys, quizzes and polls Create my own form
Privacy and cookies | Terms of use

Question 9-12:



Predicting Student Satisfaction Questionnaire

Feedback on displayed information

9. How clear was the information displayed? (1 star: not clear at all, 5 stars: very clear)

10. If you are a computer science student, what do you think about the machine learning algorithm?

Enter your answer

11. If you are a computer science student, would you have done a similar interface if you had done this project?

Enter your answer

12. If you are a computer science student, what would you have done differently?

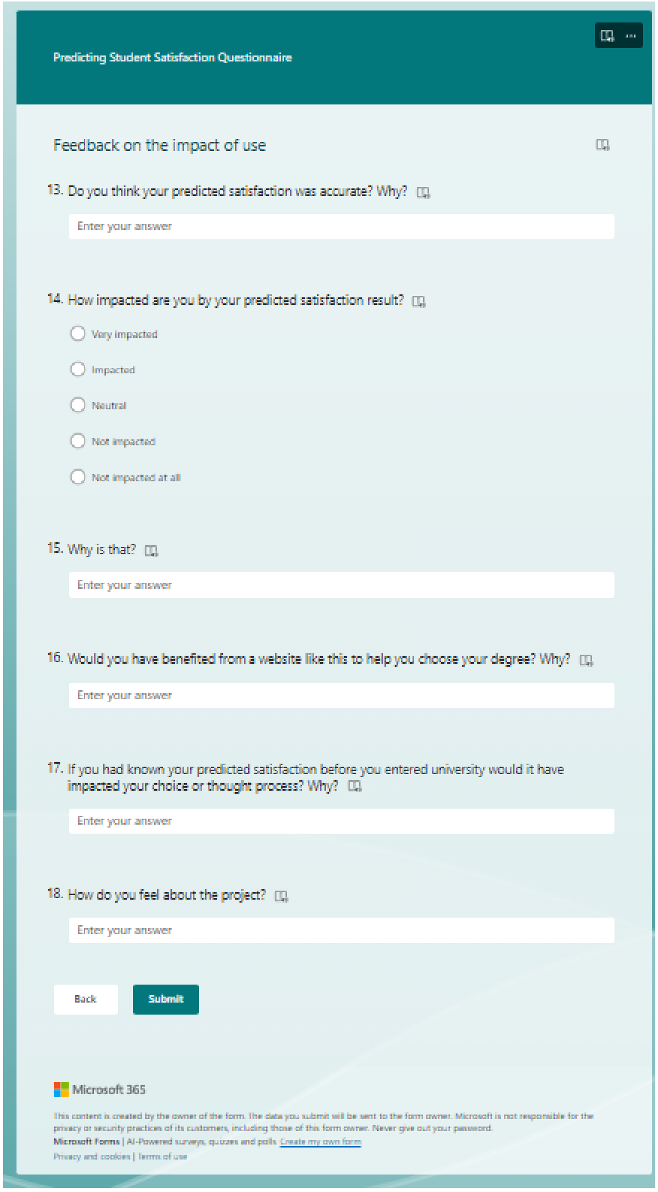
Enter your answer

Back Next

Microsoft 365

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.
Microsoft Forms | AI-Powered surveys, quizzes and polls Create my own form
Privacy and cookies | Terms of use

Question 13-18:


Predicting Student Satisfaction Questionnaire

Feedback on the impact of use

13. Do you think your predicted satisfaction was accurate? Why?

Enter your answer

14. How impacted are you by your predicted satisfaction result?

- Very impacted
- Impacted
- Neutral
- Not impacted
- Not impacted at all

15. Why is that?

Enter your answer

16. Would you have benefited from a website like this to help you choose your degree? Why?

Enter your answer

17. If you had known your predicted satisfaction before you entered university would it have impacted your choice or thought process? Why?

Enter your answer

18. How do you feel about the project?

Enter your answer

Back Submit

Microsoft 365

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Microsoft Forms | AI-Powered surveys, quizzes and polls Create my own form

Privacy and cookies | Terms of use

Appendix 4

Detailed data results for each characteristic per model.

Age	Validation Sets Post HT								
	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5	Avg	Test Set MS	MAE	R^2
LR	43.9	44.33	41.62	42.88	43.83	43.31	41.7	4.67	0.52
NB	7.38	6.01	6.9	6.83	6.63	6.75	5732.33	75.24	-64.35
KNN	31.08	31.29	29	31.74	31.08	30.84	30.76	4.07	0.65
RF	12.57	11.82	12.48	13.1	13.57	12.71	11.74	2.09	0.87
ANN						28.33	26.54	3.79	0.7
SVM	15.13	15.58	13.88	18.24	15.12	15.59	15.16	2.26	0.82
						26.16	25.18	3.38	0.71

Ethnicity	Validation Sets Post HT								
	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5	Avg	Test Set MS	MAE	R^2
LR	44.12	46.27	46.53	44.8	45.83	45.51	42.87	4.67	0.44
NB	6.09	5.68				5.89	5613.1	74.51	-72.7
KNN	32.69	36.11	33.18	33.34	33.86	33.84	33.45	3.63	0.56
RF	10.54	10.54	11.29	10.57	11.59	10.91	10.59	1.89	0.86
ANN	29.76	30.94	32.73	31.09	30.73	31.05	31.56	4.05	0.59
SVM	25.26	24.45	26.14	25.21	27.44	25.7	25.43	2.91	0.67
						29.4	28.78	3.43	0.62

Sex	Validation Sets Post HT								
	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5	Avg	Test Set MS	MAE	R^2
LR	50.04	50.86	47.04	48.46	52.87	49.85	46.76	4.79	0.43
NB	7.11	6.62	7.26	6.61	6.79	6.88	5709.58	75.09	-68.25
KNN	35.86	38.55	38.59	37.89	33.49	36.88	33.16	3.82	0.6
RF	12.66	16.36	14.46	14.91	14.33	14.54	11.82	1.99	0.86
ANN	33.43	29.83	32.44	35.09	34.21	33	31.16	3.93	0.62
SVM	27.41	28.26	25.58	26.13	28.07	27.09	24.38	2.99	0.7
						32.27	29.46	3.5	0.64

Domicile	Validation Sets Post HT								
	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5	Avg	Test Set MS	MAE	R^2
LR	39.15	42.34	40.51	39.71	43.27	41	39.51	4.48	0.44
NB	7.14	7.33	7.61			7.36	5827.28	75.96	-82.31
KNN	31.54	37.04	31.26	31.69	33.18	32.94	32.16	3.74	0.54
RF	11.98	10.95	11.66	11.77	13.14	11.9	10.48	1.84	0.85
ANN	29.28	30.09	29.54	28.75	32.75	30.08	27.83	3.66	0.6
SVM	23.24	24.17	23.77	23.93	26.96	24.41	23.62	2.9	0.66
						28.07	26.72	3.32	0.62

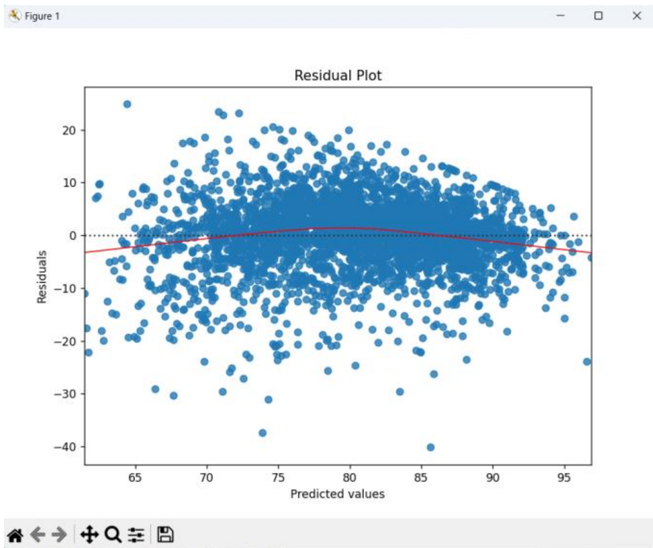
Disability Status	Validation Sets Post HT								
	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5	Avg	Test Set MS	MAE	R^2
LR	46.88	48.78	49.48	43.41	47.53	47.22	45.04	4.62	0.48
NB	7.48	8.14	7.87	7.79	8.98	8.05	5725.73	75.19	-65.64
KNN	32.61	35.26	34.29	30.34	35.1	33.52	31.46	3.96	0.63
RF	14.38	16.84	15.86	15.82	13.63	15.31	13.61	2.09	0.84
ANN	31.03	27.94	29.68	29.87	33.88	30.48	25.37	3.55	0.7
SVM	16.31	16.87	19.49	15.13	18.3	17.22	15.1	2.06	0.82
						28.75	26.12	3.26	0.69

Appendix 5

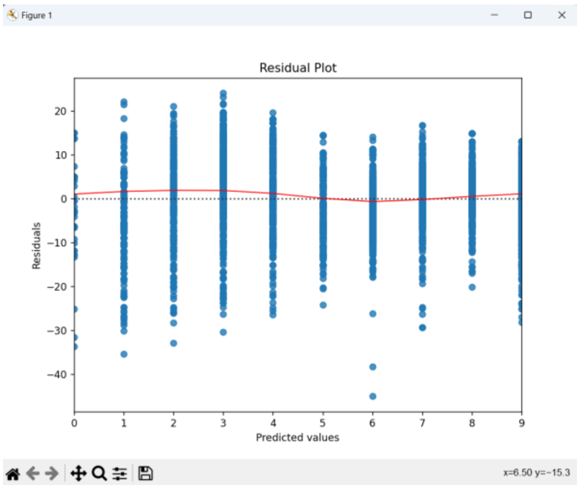
4.1 Residual Plots

4.1.1 Age plots

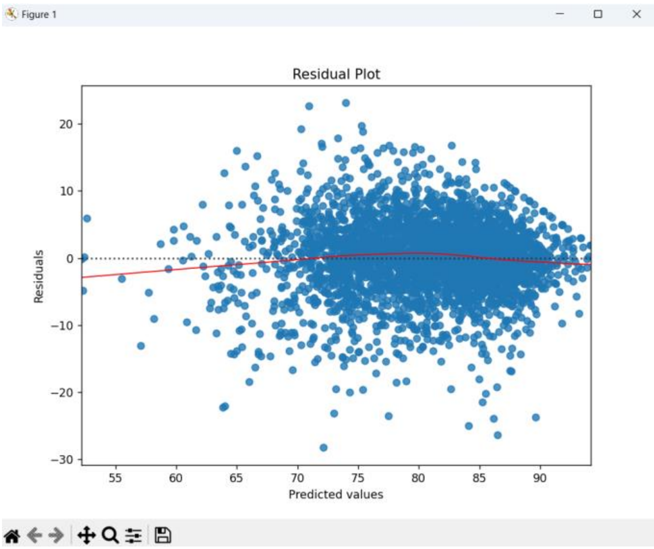
LR:



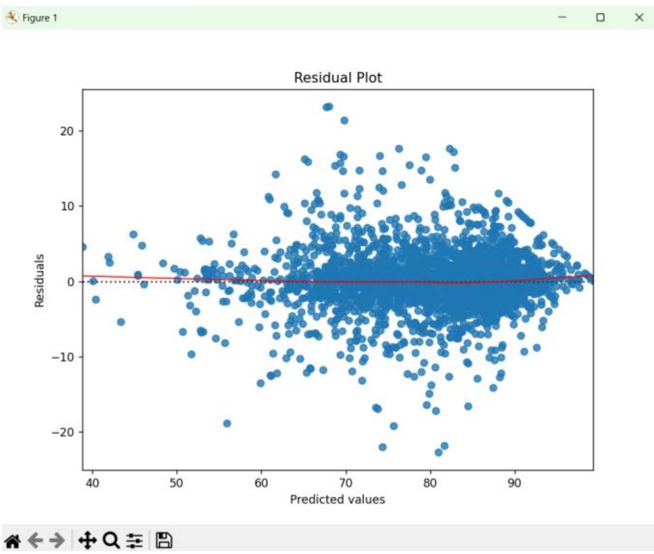
NB:



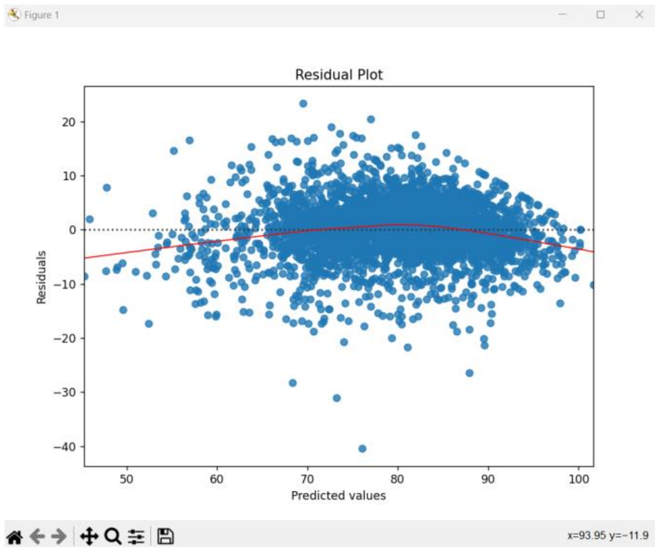
KNN:



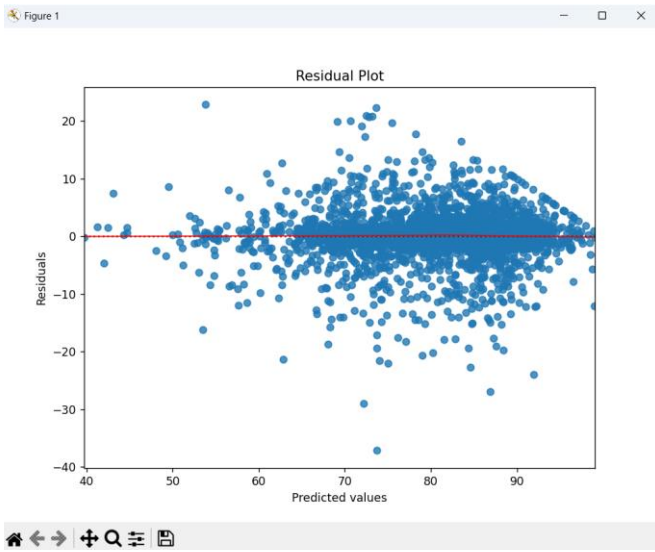
RF:



ANN:

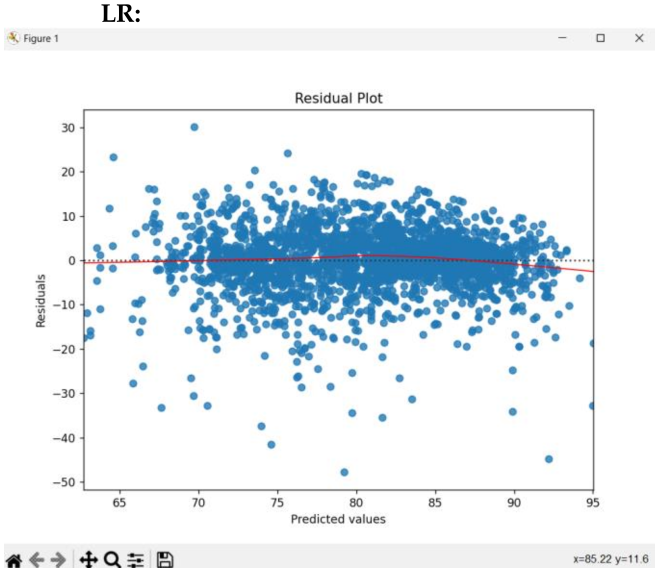


SVM:

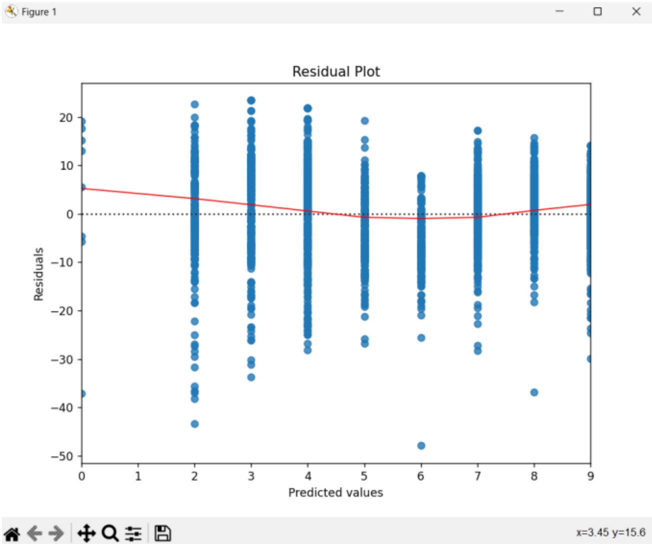


4.1.2 Sex plots

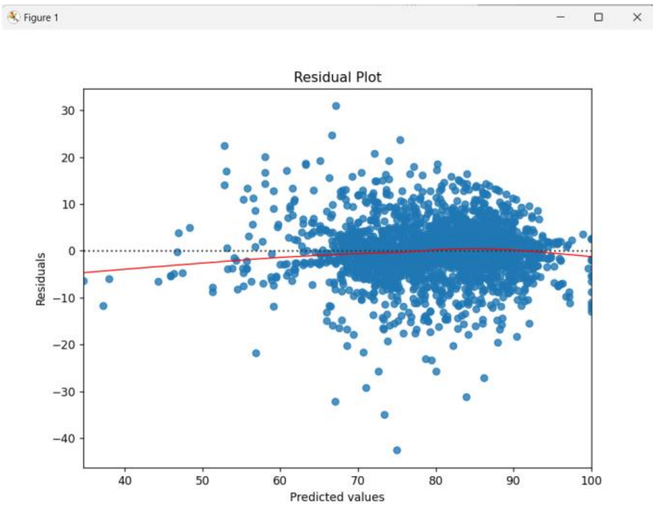
LR:



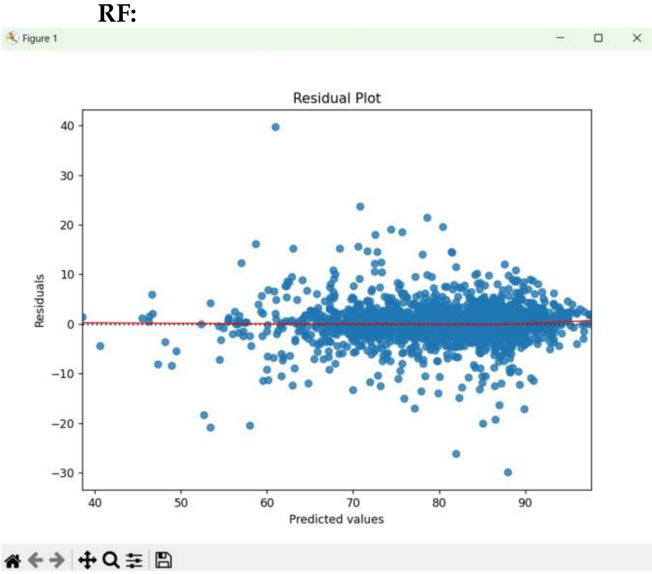
NB:



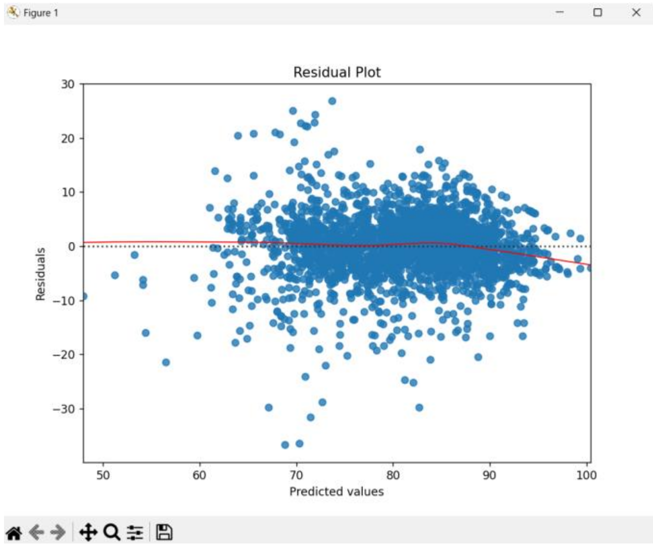
KNN:



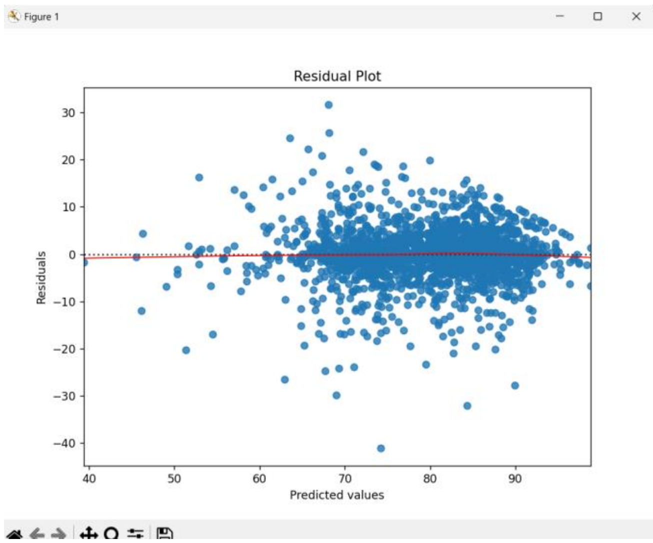
RF:



ANN:

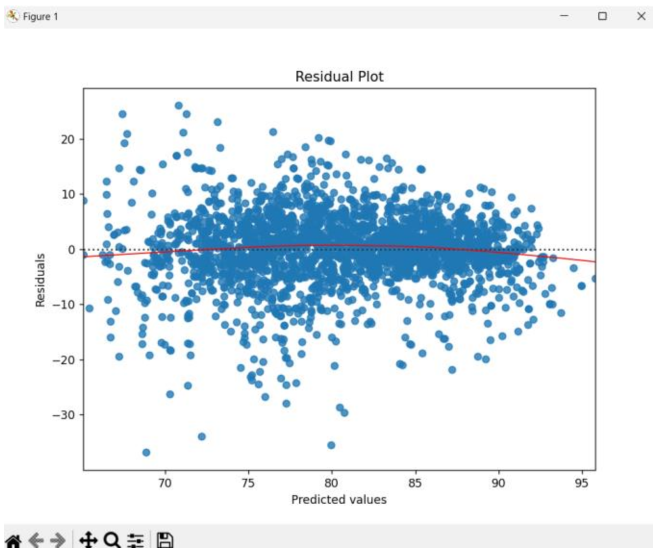


SVM:

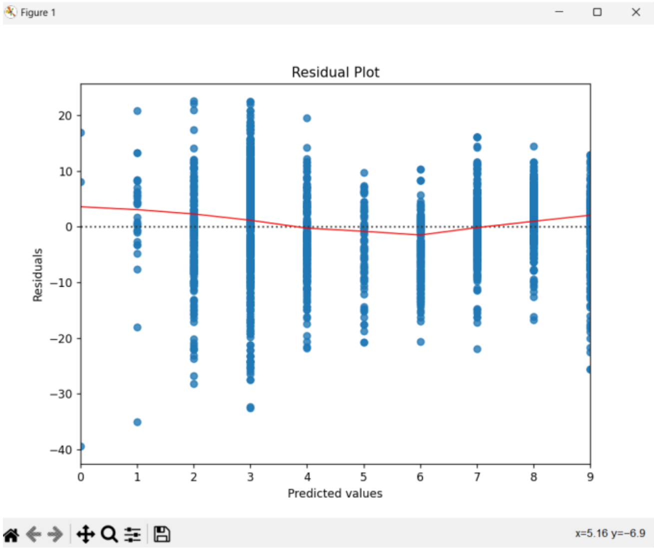


4.1.3 Domicile plots

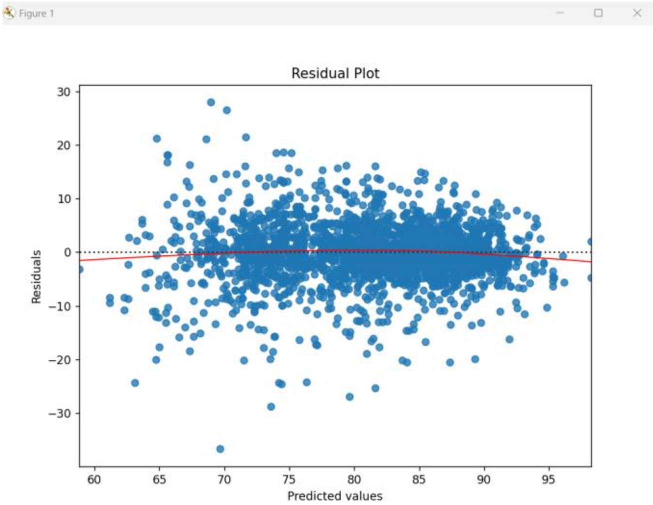
LR:



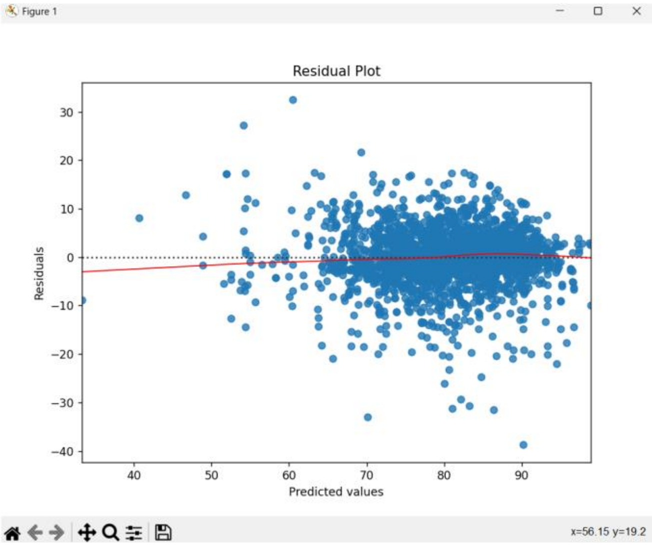
NB:



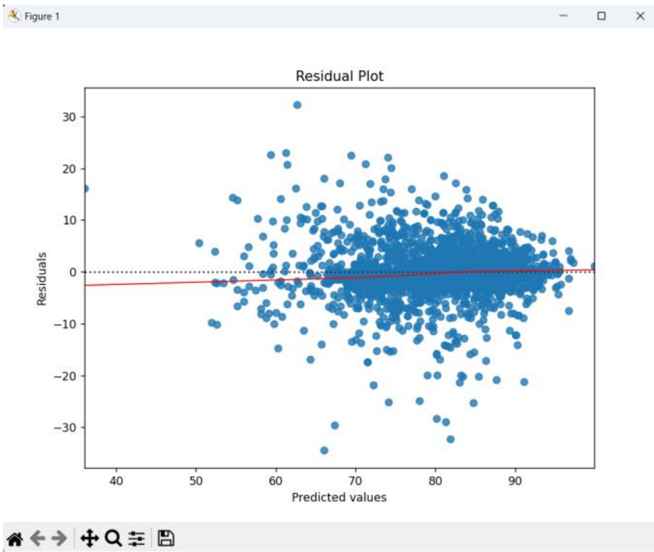
ANN:



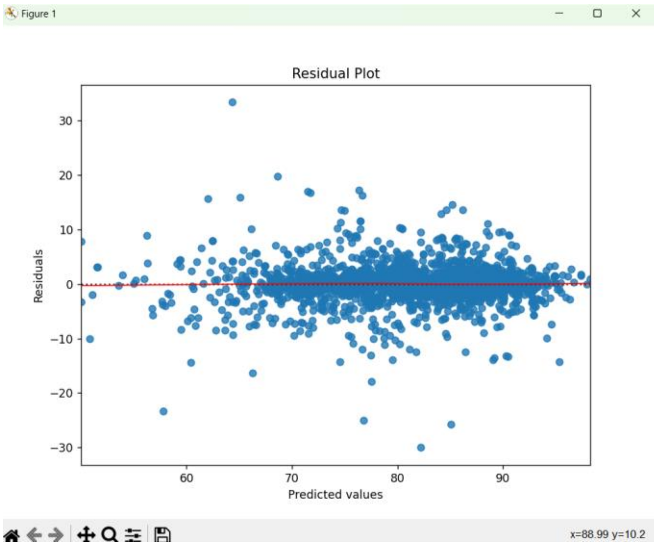
KNN:



SVM:

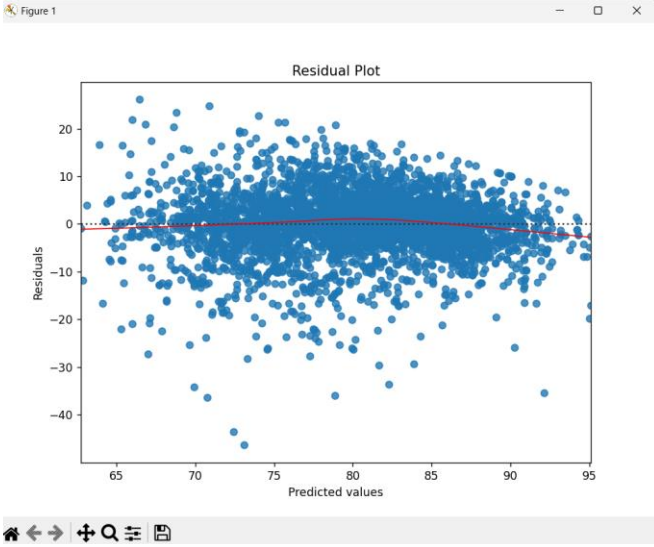


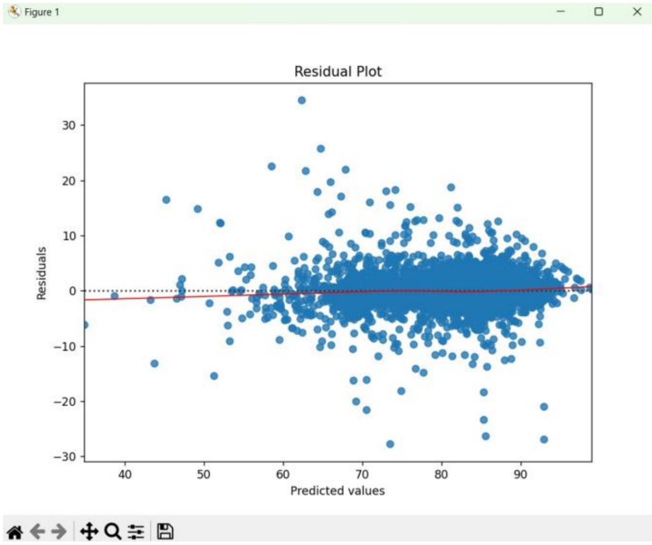
RF:



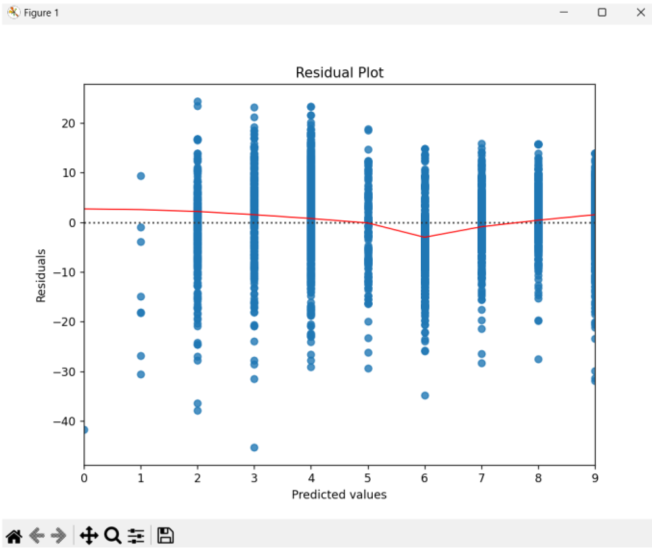
4.1.4 Ethnicity plots

LR:

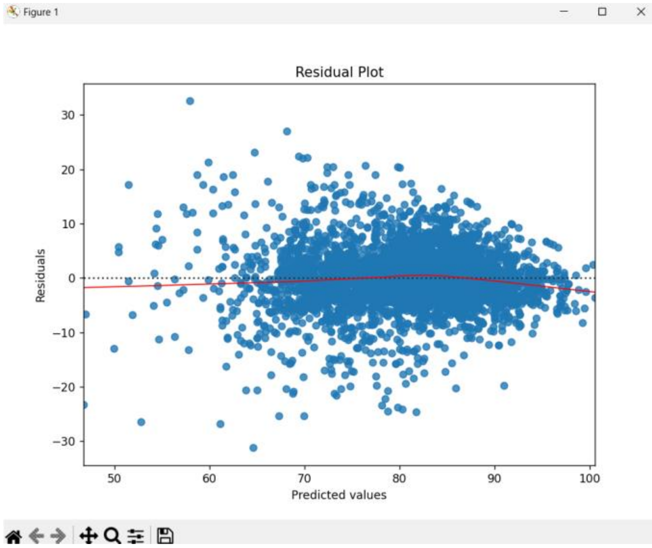




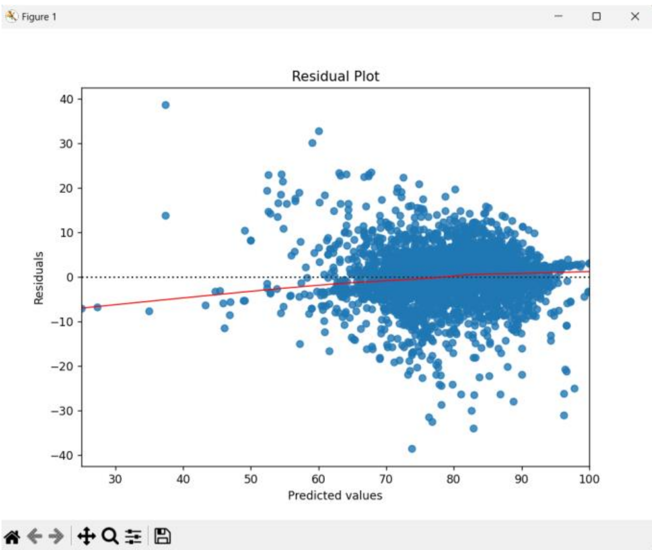
NB:



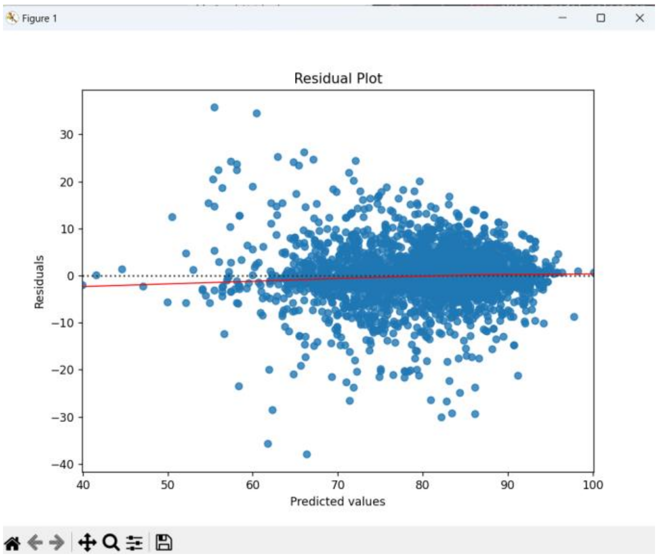
ANN:



KNN:



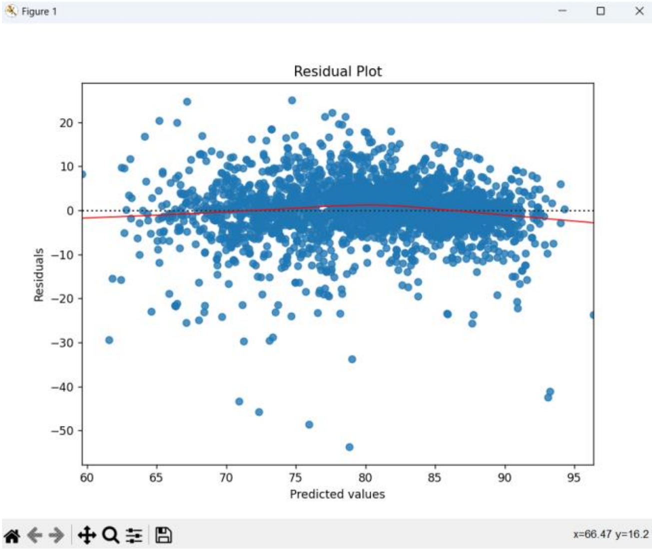
SVM:



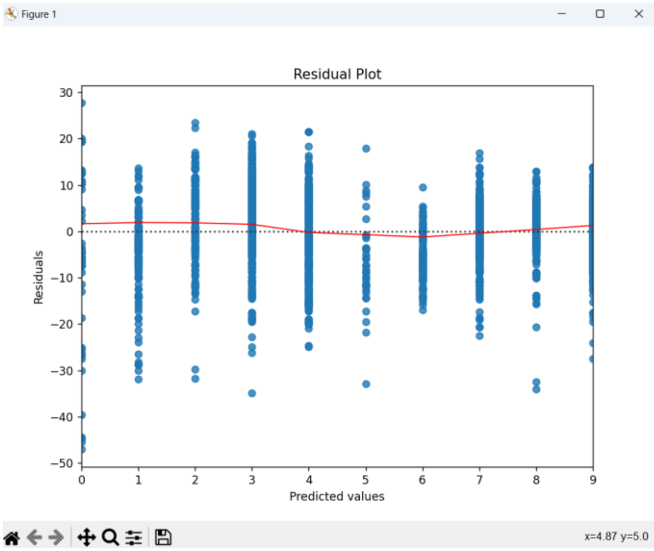
RF:

4.1.5 Disability status plots

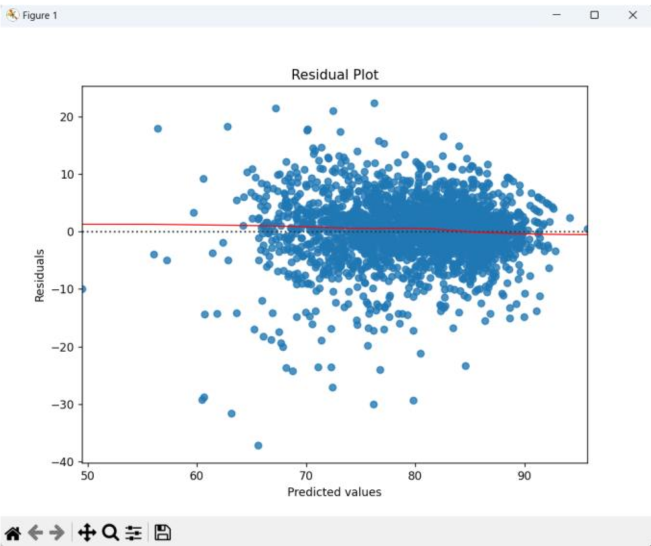
LR:



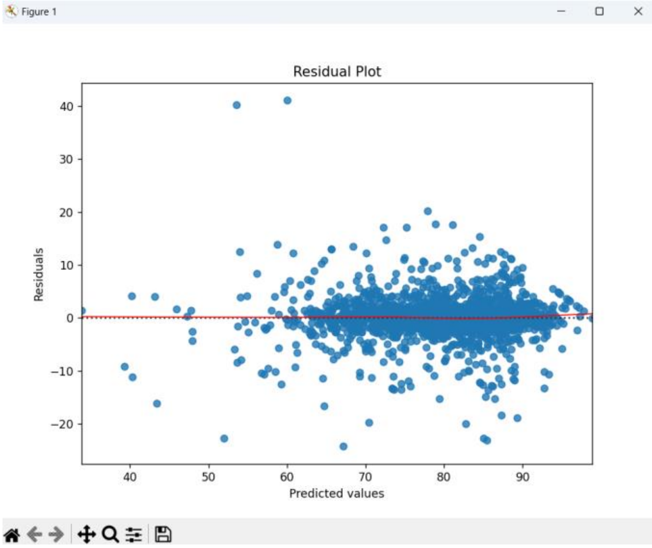
NB:



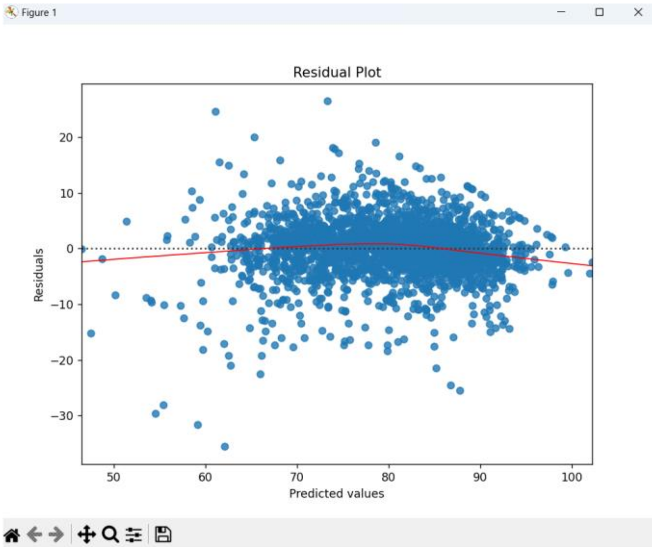
KNN:



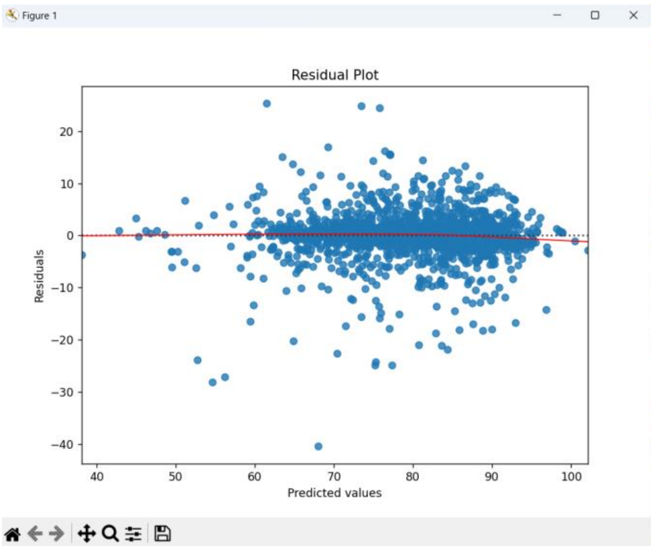
RF:



ANN:



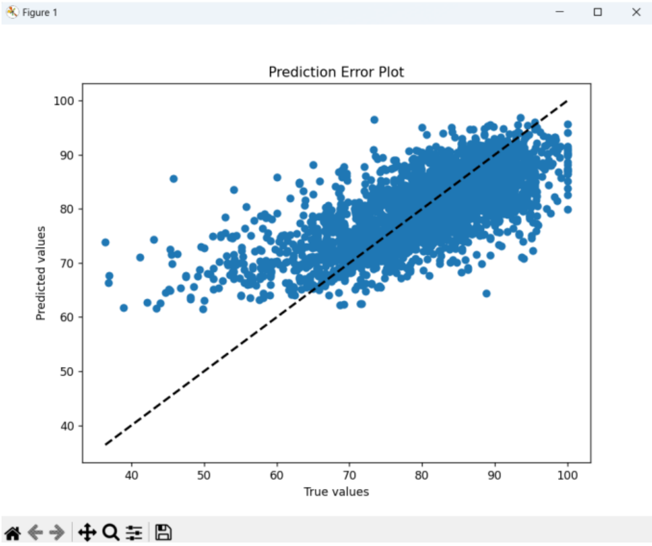
SVM:



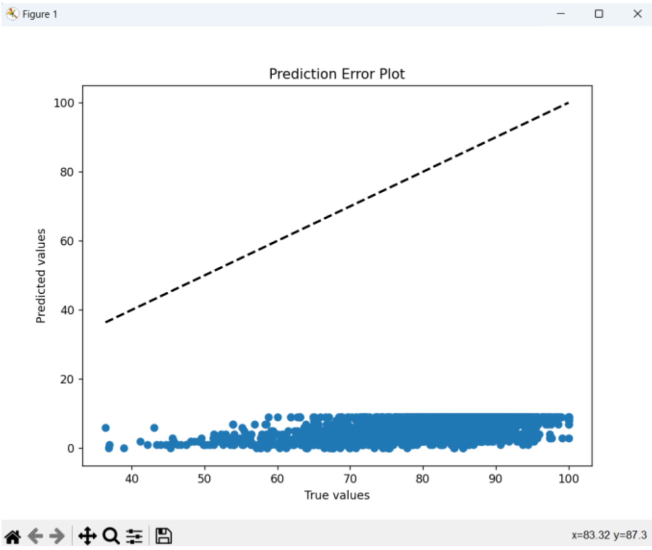
4.2 Prediction Error Plots

4.2.1 Age plots

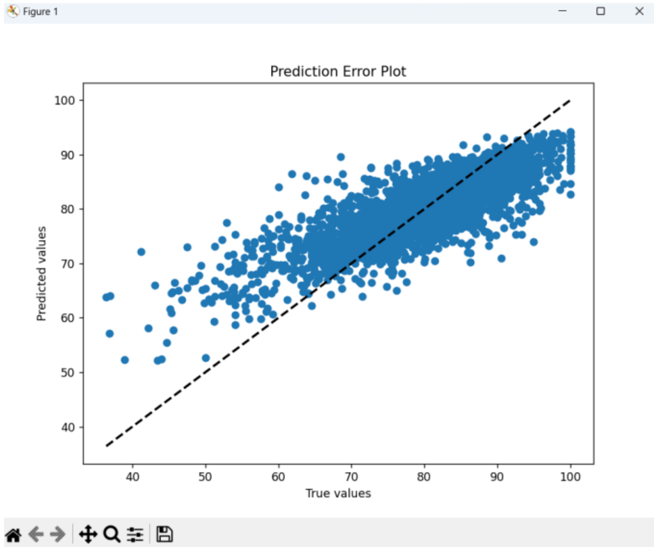
LR:



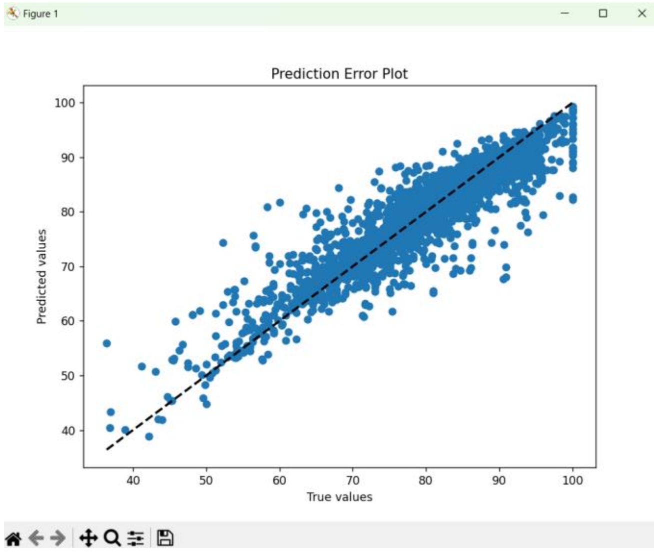
NB:



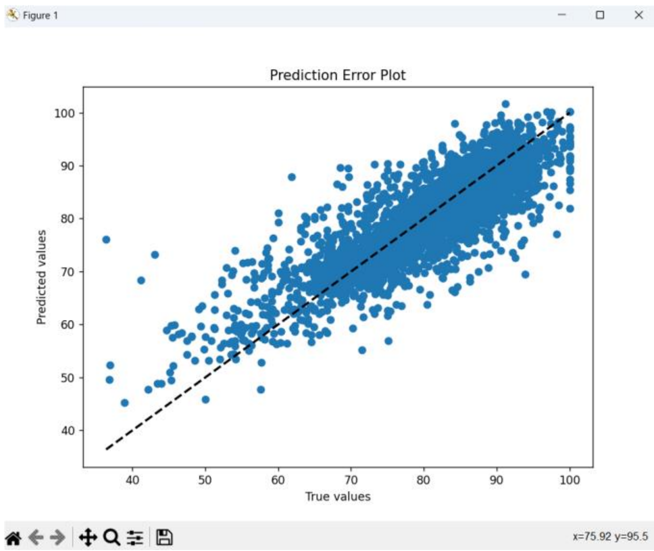
KNN:



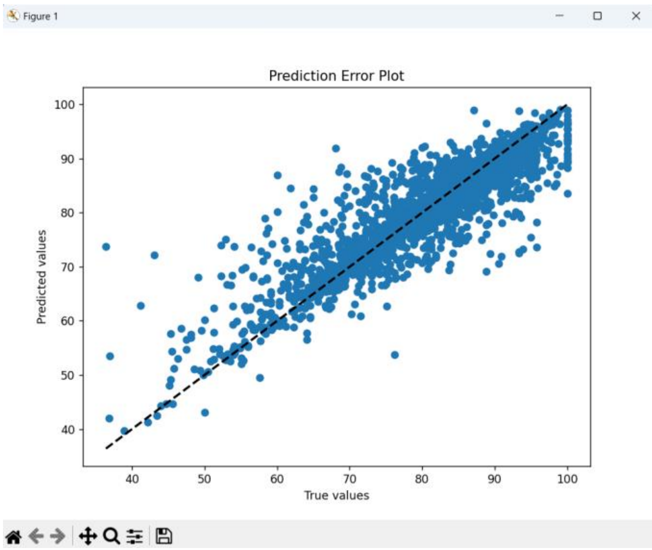
RF:



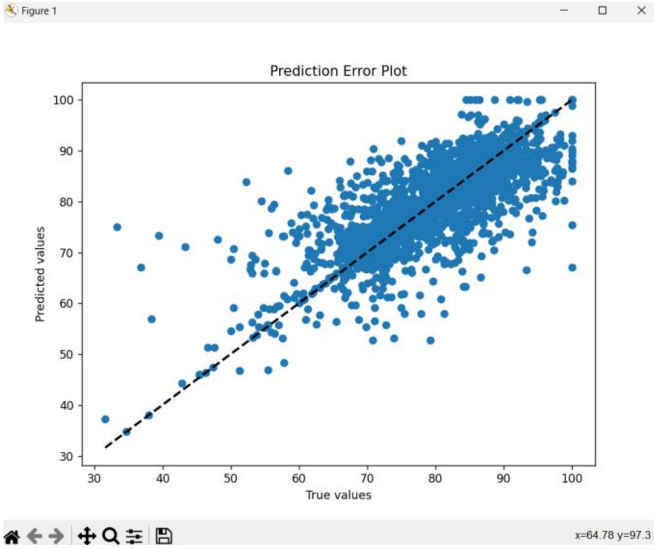
ANN:



SVM:

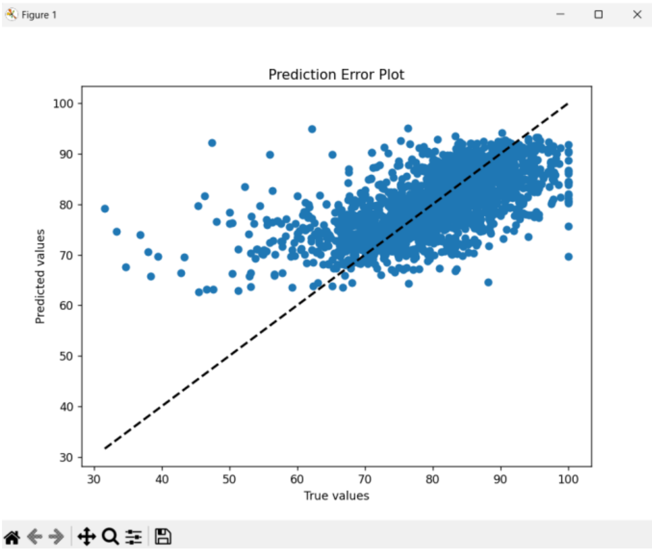


KNN:

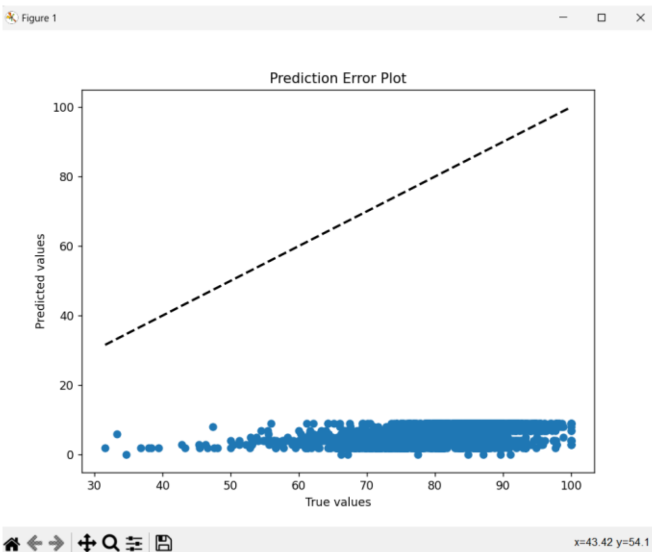


4.2.2 Sex plots

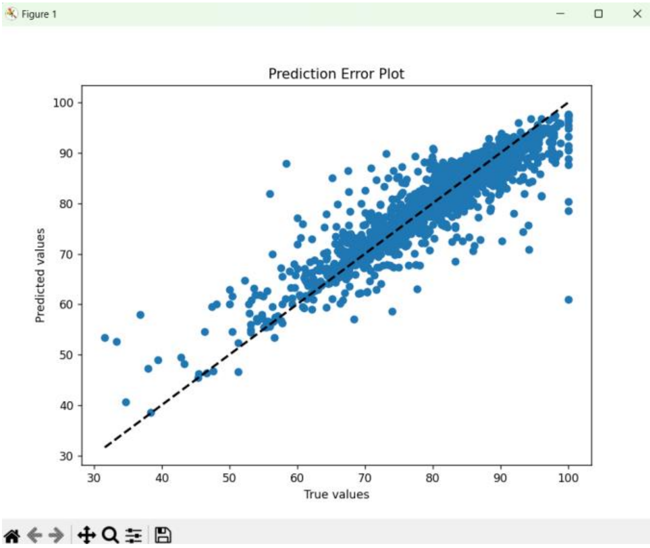
LR:



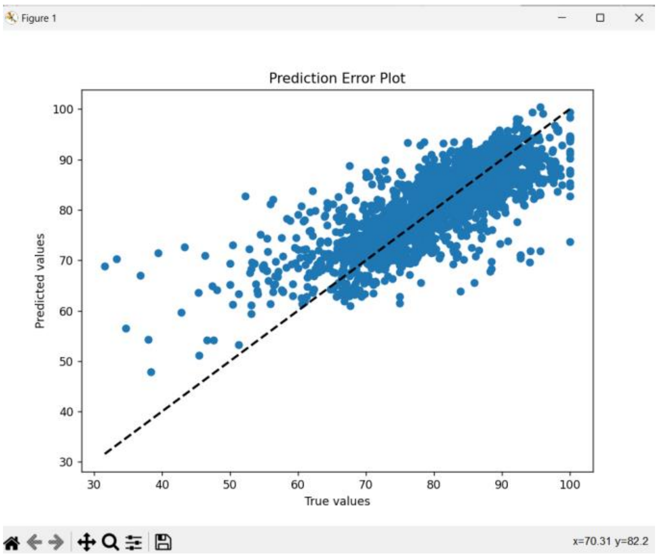
NB:



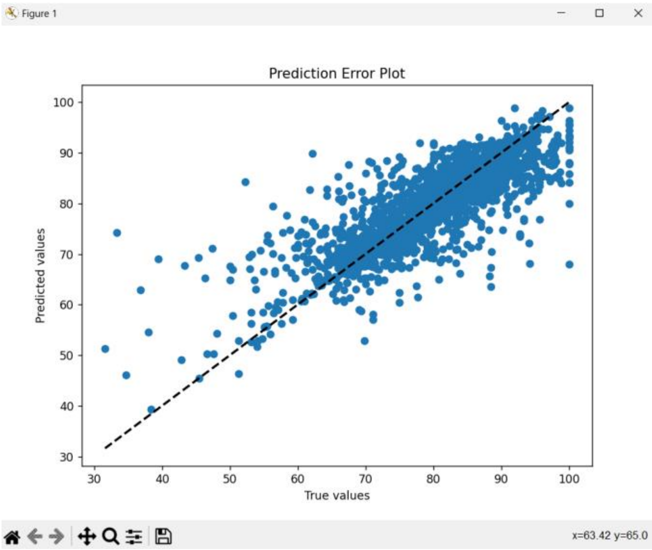
RF:



ANN:

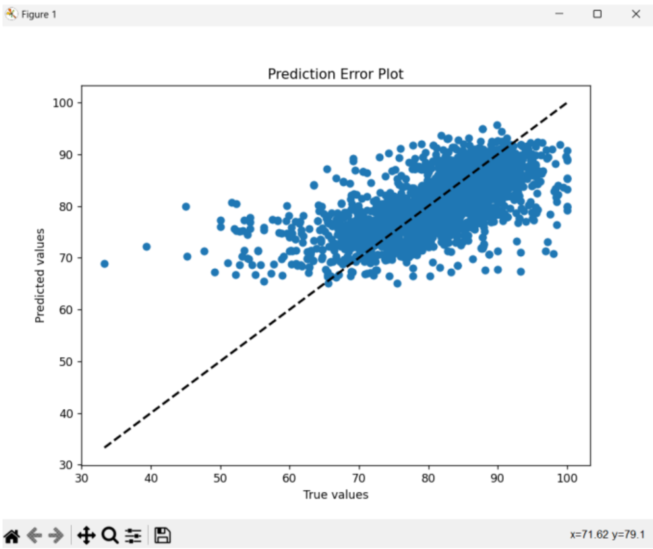


SVM:

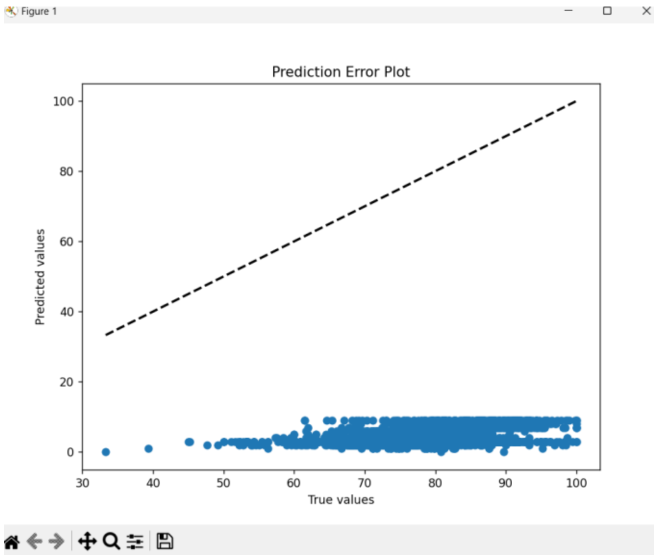


4.2.3 Domicile plots

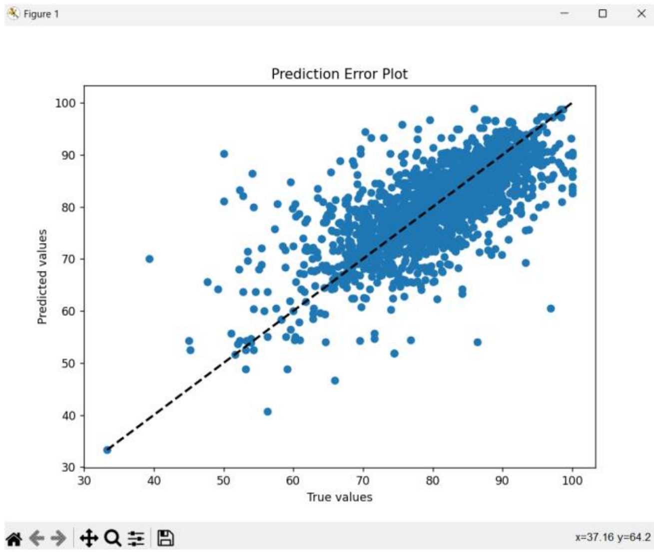
LR:



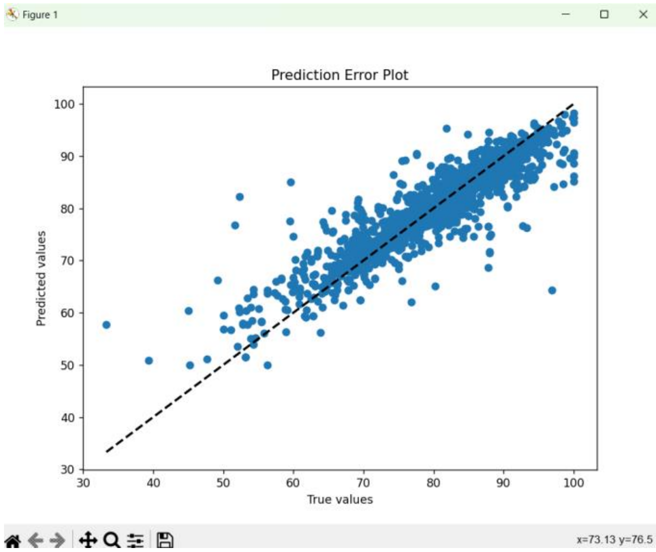
NB:



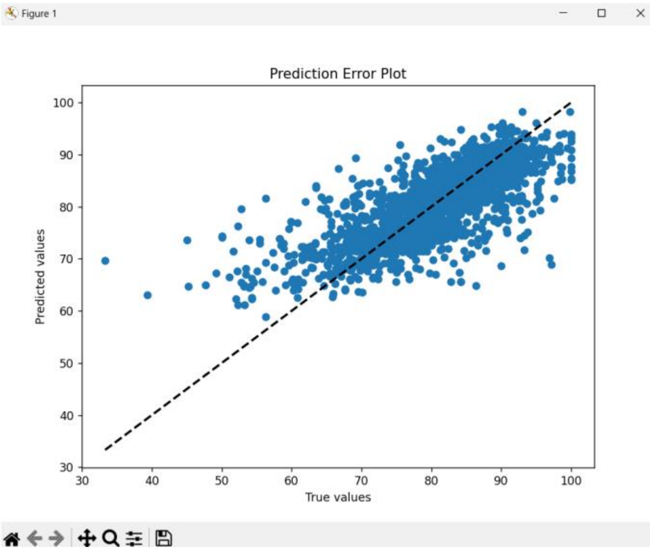
KNN:



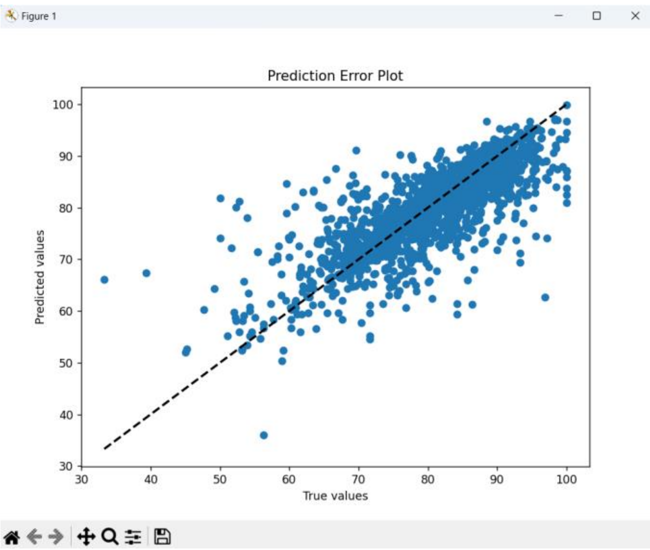
RF:



ANN:

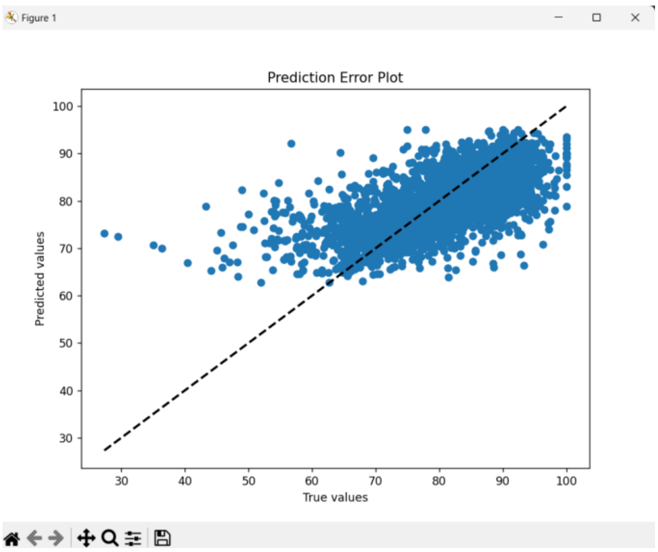


SVM:

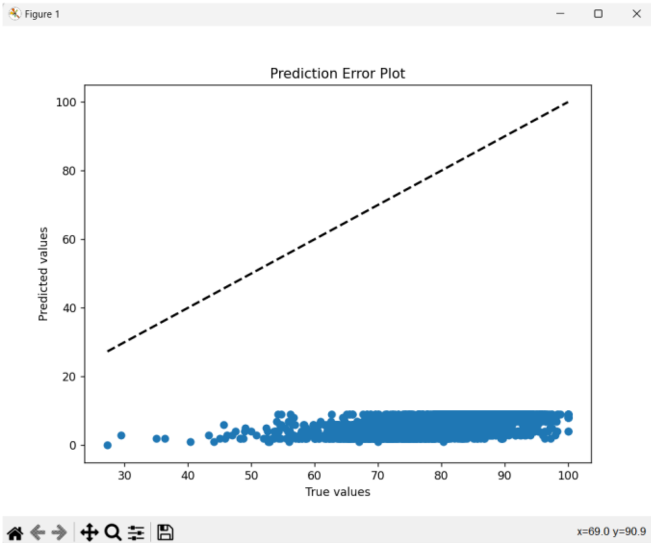


4.1.4 Ethnicity plots

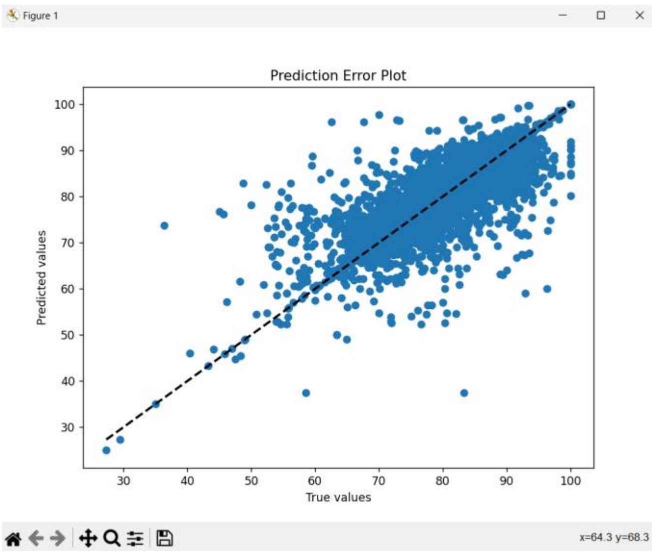
LR:



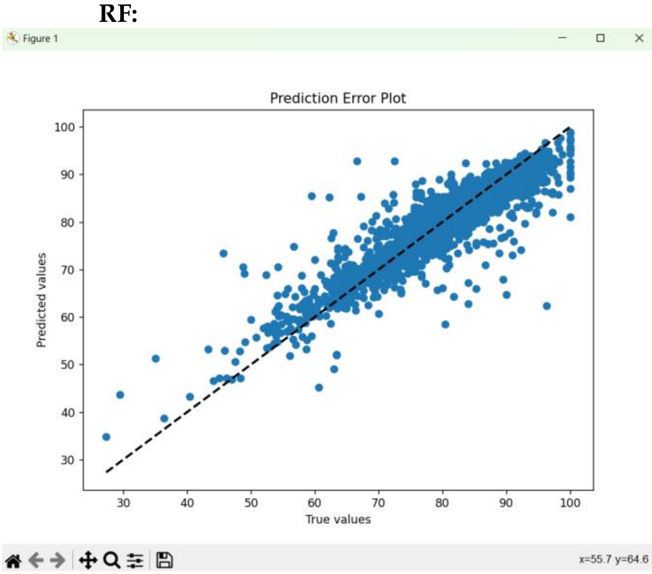
NB:



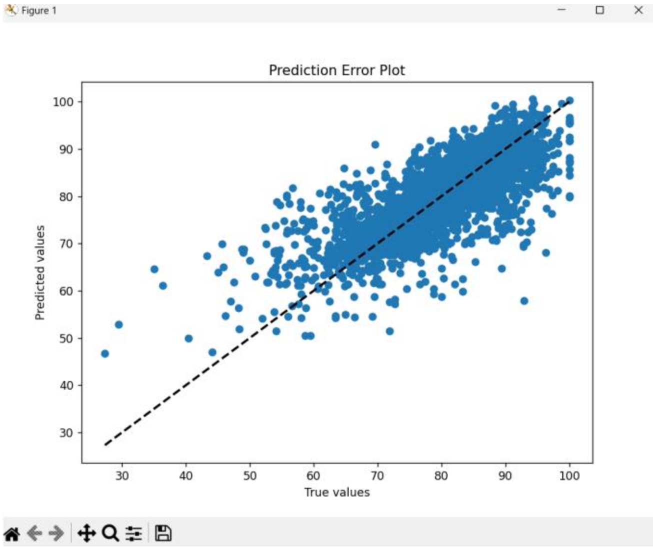
KNN:



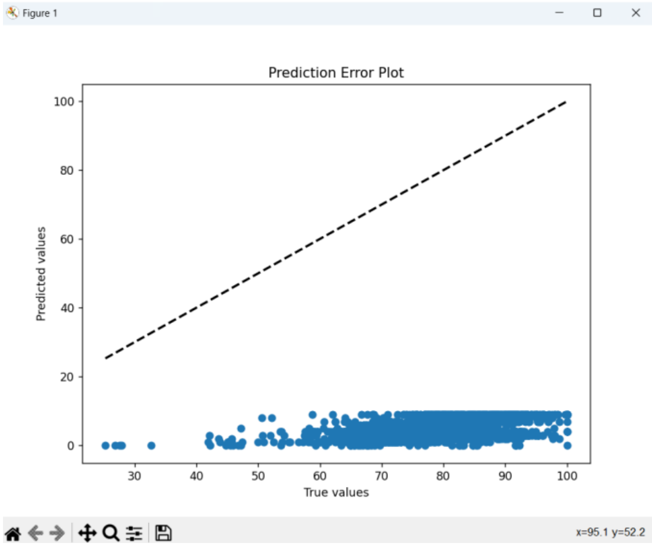
RF:



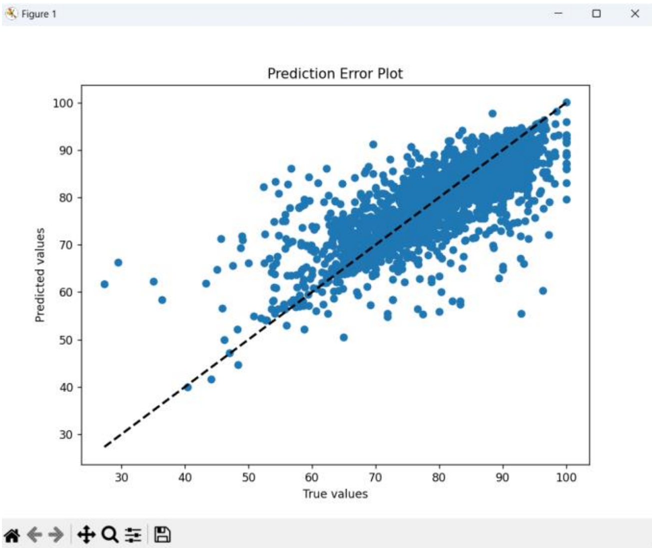
ANN:



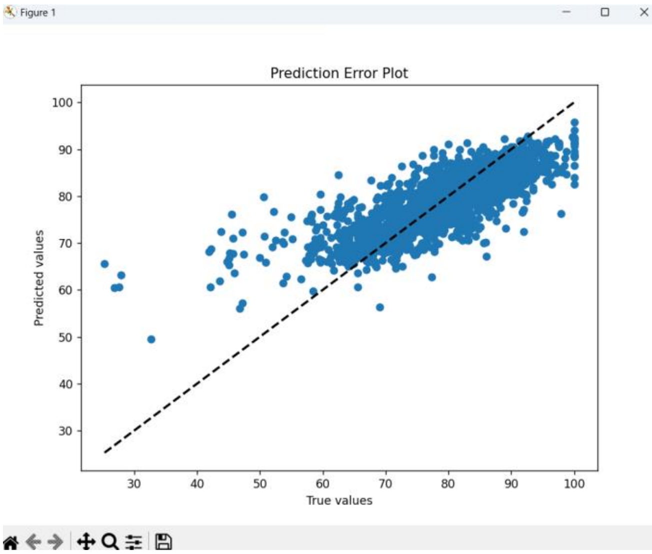
NB:



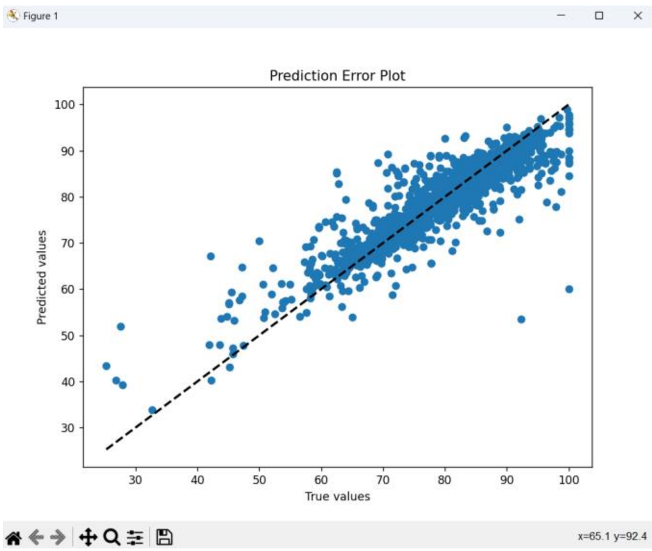
SVM:



KNN:

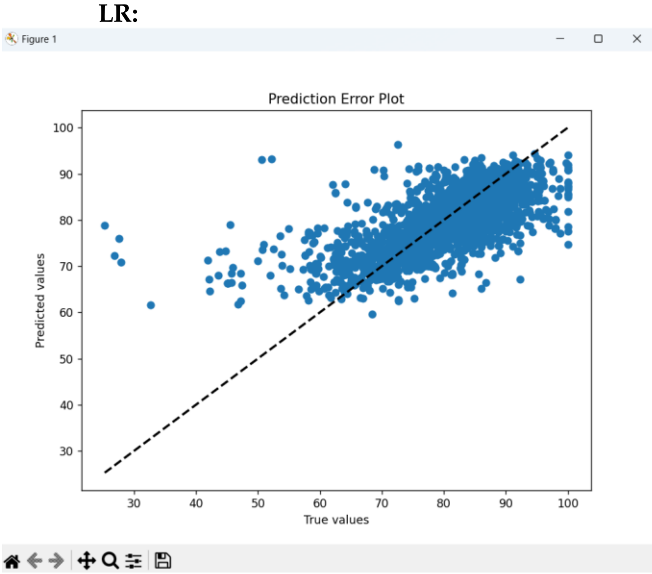


RF:

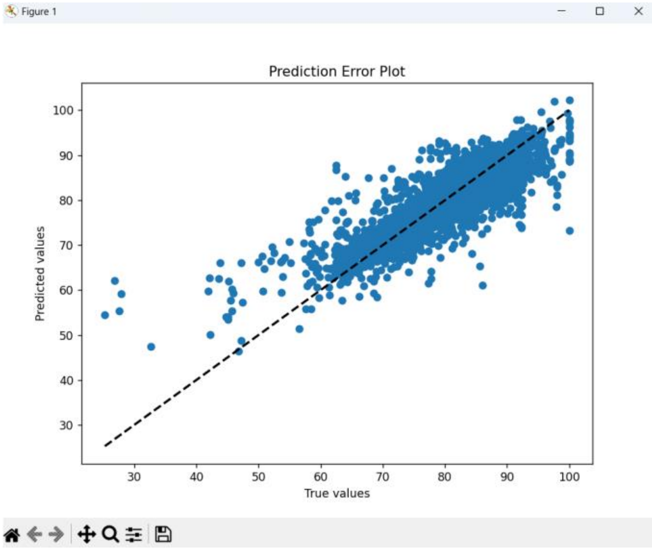


4.2.5 Disability status plots

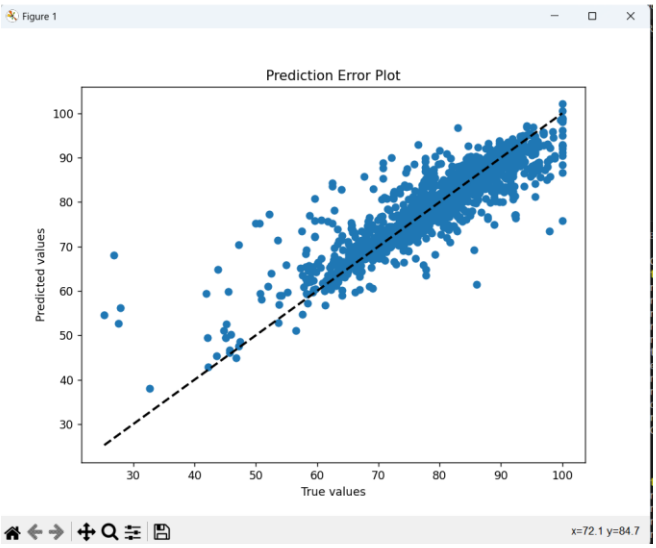
LR:



ANN:



SVM:



Appendix 6

Questions 1-5

ID	Q1	Q2	Q3	Q4	Q5
1	Yes	5	Strongly agree	Strongly Agree	Somewhat satisfied
2	Yes	5	Strongly agree	Strongly Agree	Somewhat satisfied
3	Yes	5	Strongly agree	Strongly Agree	Very satisfied
4	Yes	5	Agree	Agree	Very satisfied
5	Yes	5	Strongly agree	Strongly Agree	Very satisfied
6	Yes	5	Agree	Agree	Somewhat satisfied
7	Yes	5	Strongly agree	Strongly Agree	Very satisfied

Question 6-7

Q6

ID	Q6	Q7
1	Extremely engaging	NO

2	Very engaging	N/A
3	Very engaging	
4	Moderately engaging	no
5	Extremely engaging	n/a
6	Slightly engaging	i dont know if its an issue or just the predictions dont change but i tried predicting another characteristic and it didnt change
7	Extremely engaging	.

2	all good	5
3		5
4	The predictions are formatted in a bit of a boring and not obvious way that they were given, like it works but I hadnt realised theyd been done.	4
5	n/a	5
6	the website is basic and the layout of the predictions could be improved	5
7	.	5

Question 8-9

ID	Q8	Q9
1	NO	5

Question 10 – 12

ID	Q10	Q11	Q12
1	N/A	.	.

2			
3			
4	n/a	n/a	n/a
5	n/a	n/a	n/a
6	theyre good and usefule	yeah website are easily accessible to everyone	probably better display of the website
7			

2	Neutral	I don't actively think about satisfaction however, completing this makes me more aware of my satisfaction on the course
3	Not impacted at all	Feelings towards my course change throughout my degree but overall I'd say it was close enough
4	Not impacted	Reading the warning sign Im aware that this isnt accurate
5	Not impacted at all	because this is still a prototype so im not really worried by its prediction
6	Not impacted	cus its not accurate and we were warned so thats fine
7	Not impacted	not really that bothered

Question 13 – 14

ID	Q13	Q14
1	Impacted	Made me appreciate my course, motivate me to be more proactive with my as signments.

Question 15

ID	Q15
1	YES, I'd want to know whether I'd enjoy a future course but also I'm interested in using tools and websites to give me a better idea about courses and future working opportunities before going into them.

2	No, my choice in degree was led by interest not satisfaction
3	No due to me always being set on my course from a young age
4	Yes I would have chosen another degree had I known about my satisfaction rate with my current degree
5	yes, satisfaction i think impacts experience greatly and if i knew where i would be satisfied i would have prioritised those degrees more
6	yes it would help better choose a degree because we predict our satisfaction based on how satisfied we are with that subject in school but school and uni are two very diff things
7	yeah but also no like satisfaction wasnt my top criteria to pick a degree

Question 16

ID	Q16
1	Maybe' because I did travel from home to be at Uni, when I could've stayed more local.

2	potentially, it would allow me to find out where i would be best satisfied doing my degree which would factor into decision making
3	Potentially between universities but not between different courses in the same uni
4	Yes definitely I would have thought through more about my degree choice and probably chosen a different one rather than picking my fave subject
5	yes, what i said in q16
6	yeah i would take a closer look at different degrees rather than favouritising this one
7	yeah i would have maybe taken it into account

Question 17

ID	Q17
1	Easy to follow, clear and useful.

2	this is a good accessible way to determine satisfaction which may be very beneficial for secondary school students to assist them in making a choice between universities or whether to actually go to university or not
3	I think it could be useful for future students for final decisions
4	I think this is a really good project and would be very valuable if expanded and made accurate for future potential student
5	its good
6	its a decent project and if the student had longer they could really expand on it and make it into something that could be used in the world
7	its good

2	Fairly accurate, roughly close to how i would rate it
3	Yes; overall satisfaction and teaching is similar to how i feel
4	No not really, I've not been satisfied with my course at university and the way it was delivered however it said I was
5	kind of, im satisfied in some aspects and it was correct there however im unsatisfied for some themes
6	no im not satisfied with certain aspects of my course like organisation and management theme 5 but it predicted 71.9% satisfaction
7	yeah im satisfied with my course

Question 18

ID	Q18
1	I am pretty sqtisfied with my course anyway so things seemed accurate to me. It did highlight areas that might be better though, which was understandable.

Appendix 7

Below is the table with all the hyperparameter comparison data of pre vs post for each model and characteristic.

The table is unable to shift into central position.

	ANN		KNN		LR		NB		RF		SVM	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Age	41.61	28.33	42.16	30.84	43.72	43.71	9.63	6.94	12.86	12.88	33.42	15.39
Sex	48.77	33	40.05	36.88	49.76	49.85	8.59	6.88	14.38	27.2	41.38	27.09
Domicile	39.52	30.08	33.52	32.94	40.71	41.2	7.95	7.36	12.38	23.94	31.61	24.21
Ethnicity	42.93	31.05	35.56	33.83	45.3	45.51	6.33	5.88	10.89	22.98	34.96	25.7
Disability status	46.58	30.48	39.12	33.52	47.04	47.02	11.15	8.05	15.44	15.37	39.16	17.22
Average	43.882	30.588	38.082	33.602	45.306	45.458	8.73	7.022	13.19	20.474	36.106	21.922