Tutor review score prediction

Elizaveta Antonenko

2023

Abstract

The work is devoted to predicting the evaluation of reviews about tutors in Moscow and St. Petersburg. It is very important to be able to evaluate reviews about services for statistics and for cases when user did not rate or gave an incorrect rating. Here you will find complete information about my work: `https://github.com/Elizaveta-Pavlovna/Predict-review-score`.

1 Introduction

Social networks and services are constantly evolving, and more and more people can express their opinion about goods and services publicly. To determine the attitude of users to something, you have to analyze tens of thousands of messages. It becomes impossible to process such a volume of information manually. The work analyzes data with feedback about tutors to determine the review score (one of four: 2, 3, 4, 5)

1.1 Team

Antonenko Elizaveta analyzed the data and prepared this document.

2 Related Work

The task of classifying reviews is very popular in the world of data analysis, but no solutions were found specifically on the topic of analyzing reviews of tutors and, of course, on this data, because they were collected specifically for this work.

3 Model Description

The task is to apply and compare well-known classification algorithms, presented in the table:

Algorithm	Main idea
Logistic Regression	Calculating the probability that a given input value belongs to a particular class
Decision Tree	Method of representing decision rules in a hierarchical structure
Random Forest	Finding the best, among a large ensemble of decision trees
Gradient Boosting	Building a prediction as an ensemble of weak predictive models
Ada Boost	Building a linear combination of classifiers
Complement NB	A special case of Naive Bayes classifier

4 Dataset

For analysis, using page parsing, the dataset of user reviews of the site `https://spb.repetitors.info/repetitor/` about the services of teachers (and in very rare cases, representatives of other fields of activity, such as a hairdresser, recruiter, animal nurse) is collected. The following cities were considered: Moscow, St. Petersburg, Kazan, Veliky Novgorod, Yekaterinburg, Krasnodar, Rostov-on-Don. Further, the data was combined into one table with 132689 rows and 2 columns "comment" and "mark".

	comment	mark
0	Преподаватель находит индивидуальный подход п...	Оценка: 5+.
1	Плюсы: Анна Валерьевна обязательна компетентн...	Оценка: 5.
2	СПАСИБО Анна Валерьевна! Сдали хорошо ЕГЭ. С ...	Оценка: 5.
3	50 % успеха - наличие контакта с преподавател...	Оценка: 5+.
4	Анна Валерьевна является не только уникальным...	Оценка: 5+.
...	...	...
7223	С Марией Николаевной мы занимаемся не так час...	Оценка: 5.
7224	Хорошо. Она хорошо преподает. Понятно доступ...	Оценка: 5.
7225	В принципе все положительно. Замечательный п...	Оценка: 5.
7226	Мы занимаемся до сих пор. Впечатления положит...	Оценка: 5.
7227	Работа репетитора понравилась замечаний нет....	Оценка: 5.

132689 rows × 2 columns

During preprocessing, a large amount of data that affects the imbalance of classes was removed (classes with grades 5 and 5+ outweighed all others by 15 times). Further, all reviews were divided into tokens, brought to the initial form. And also removed stop words and reviews, less than 7 words long (since during manual analysis it was found that short reviews do not contain much information and they are mostly neutral)

	Train	Test
Articles	14618	6266

Table 1: Dataset statistic.

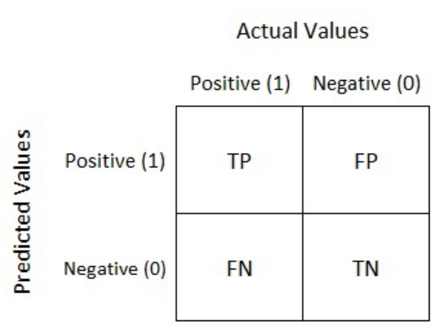
5 Experiments

Work plan:

1. data collection,
2. preprocessed block,
3. vectorization data by TF-IDF method,
4. split all the data into test and train sets,
5. training models
6. calculation of quality indicators
7. comparison of built models

5.1 Metrics

Use the error matrix:



The following quality metrics were calculated in the work:

1. **Precision** - the proportion of objects called positive by the classifier and at the same time really being positive: $\frac{TP}{TP+FP}$

2. **Recall** - shows what proportion of objects of a positive class out of all objects of a positive class was found by the algorithm: $\frac{TP}{TP+FN}$

3. **F1-score** - harmonic mean of precision and recall: $\frac{2(Precision*Recall)}{Presicion+Recall}$

5.2 Experiment Setup

After preprocessing, the data were divided into test and train samples in the ratio of 30:70. Next, a function "model_score" was implemented that received a classifier and a quality criterion as input. In this function, the input model was trained on the training data and predicted the output values for the test set. The output of the function was the value of the selected quality criterion.

Firstly, models were built with default hyperparameters. Then, with GridSearchCV, the best parameters for some models were found.

5.3 Baselines

In our case using TF-IDF with linear regression as baselines.

6 Results

Model name	**Precision**	**Recall**	**F1-score**
LogisticRegression	0.807458	0.760294	0.777502
RandomForestClassifier	0.865081	0.715608	0.766449
ComplementNB	0.845159	0.719438	0.763849
DecisionTreeClassifier	0.639376	0.635653	0.635638
GradientBoostingClassifier	0.799436	0.726141	0.753830
AdaBoostClassifier	0.757067	0.698053	0.720645

Table 2: Model results.

Model name	**Best params**	**Precision**	**Recall**	**F1-score**
LogisticRegression	'C': 1.0, 'tol': 0.0001	0.807458	0.760294	0.777502
RandomForestClassifier	'max_depth': 4, 'max_features': 3, 'min_samples_split': 3, 'n_estimators': 20	1.0	0.627673	0.771252
ComplementNB	'alpha': 0.1	0.72431	0.711459	0.712507

Table 3: Model with Grid params results.

With GridSearchCV models got lower results than with default parameters.

As a result, RandomForestClassifier and ComplementNB from Table 2 provided the best results.

Examples of how the model works for reviews with different ratings:

Comment	Real Mark	Pred mark
Рекомендую! Ирина Владимировна-очень грамотный педагог.Доступно и понятно объясняет материал. Дочери очень нравится.	5	5
Занимались до конца учебы. Сдал сын не очень хорошо но в этом нет вины Артура Валерьевича. Я считаю что в этом вина ребенка что он расслабился. У них хорошо шли занятия хороший был контакт. Хорошо что по этому предмету сын с молодым человеком занимался легко ему было из-за этого. Сдал сын ЕГЭ на 40 баллов.	4	4
Занятия прошли без особых эмоций ожидал большего. Объяснения были сухи и скучны. Хотелось бы от репетитора больше заинтересованности.	3	3
Не рекомендую данного человека называющего себя репетитором для подготовки подростков к ОГЭ и ЕГЭ. К сожалению довольно длительно (в течение почти года) занимались но к счастью сложившиеся обстоятельства нас развели . И после тестирования с настоящим педагогом русского языка узнали о потерянном времени и немалых деньгах. Очень жаль что нет договора об услугах которые продает данный человек. Не рекомендую даже предостерегаю!	2	2

Table 4: work examples.

The dataset also contained sarcastic comments that the model could not recognize. For example:

'ИЛЬМИРА ДАМИРОВНА ПРОВЕЛА ЗАНЯТИЯ ОГРОМНОЕ СПАСИБО! ЗАНИМАТЬСЯ НАЧАЛИ С ДРУГОЙ ДАТЫ А НИ ТОГДА КОГДА БЫЛО ЗАЯВЛЕНО ПРИ ПОИСКЕ УЧИТЕЛЯ. ГРАФИК ЗАНЯТИЙ ЧАСТО МЕНЯЛСЯ

СЧИТАЮ ЧТО ПЕДАГОГ НЕ СОВСЕМ КОРРЕКТНО ПОДОБРАН НАМ НУЖНО БЫЛО ПОВТОРИТЬ ШКОЛЬНУЮ ПРОГРАММУ НАЧАЛЬНЫХ КЛАССОВ И ПОДГОТОВИТЬ РЕБЕНКА К 5 КЛАССУ. ИЛЬМИРА ДАМИРОВНА ПЕДАГОГ СТАРШИХ КЛАССОВ СООТВЕТСТВЕННО МЕТОДИКА РАБОТЫ ОТЛИЧАЕТСЯ....'.

The comment was rated 3 by the user and 5 by the model. Most likely influenced by "огромное спасибо".

7 Conclusion

As a result, we collected a large dataset for the project, split the data into 4 classes, performed preprocessing, trained different models and compared them. From which we made a point analysis of the results of the best model.