

## Сравнение ансамблей методов машинного обучения для прогнозирования оттока клиентов

*Гурков Игорь Александрович*

*Морской государственный университет им. адм. Г.И. Невельского*  
*Студент*

*Волков Никита Михайлович*

*Морской государственный университет им. адм. Г.И. Невельского*  
*Студент*

### Аннотация

В работе осуществлен анализ ансамблей методов машинного обучения для прогнозирования оттока клиентов. Предложены наилучшие ансамбли для решения задачи оттока клиентов.

**Ключевые слова:** машинное обучение, отток клиентов, ансамбль, бустинг, анализ данных.

## Comparison of ensembles of machine learning methods for predicting customer churn

*Gurkov Igor Alexandrovich*

*Maritime State University adm. G.I. Nevelsky*  
*Student*

*Volkov Nikita Mikhailovich*

*Maritime State University adm. G.I. Nevelsky*  
*Student*

### Abstract

The paper analyzes the ensembles of machine learning methods to predict the outflow of customers. The best ensembles for solving the problem of customer churn are suggested.

**Keywords:** machine learning, customer churn, ensemble, boosting, data analysis.

Набор данных IBM содержит информацию о клиентах телекоммуникационной компании и о том, покинули ли они компанию в течение последнего месяца (отток). Каждая строка представляет уникального клиента, в то время как столбцы содержат информацию об услугах клиента, учетной записи и демографические данные. Набор данных состоит из 21 столбца и 7043 строки [1].

Каждый столбец представляет собой характеристику (признак) для обучения, к таким признакам относятся: customerID (Уникальный идентификатор клиента), gender (пол), SeniorCitizen (пожилой гражданин),

Partner (партнер), Dependents (наличие иждивенцев), tenure (срок пользования), PhoneService (наличие телефона), MultipleLines (многоканальность), InternetService (Интернет служба), OnlineSecurity (Интернет-безопасность), OnlineBackup (Резервные копии онлайн), DeviceProtection (Защита устройства), TechSupport (Техническая защита), StreamingTV (Прямые трансляции телевидения), StreamingMovies (Прямые трансляции фильмов), Contract (Тип заключенного договора оплаты), PaperlessBilling (Безбумажное выставление счетов), PaymentMethod (Тип оплаты клиента), MonthlyCharges (Месячная плата), TotalCharges (Сумма плат), Churn (Отток).

Для работы с данными воспользуемся библиотеками для анализа данных ЯП Python: pandas, sklearn.

Во время ансамблирования моделей машинного обучения было использовано 9 методов машинного обучения, а именно: наивный байесовский классификатор (Naive Bayes), случайные деревья (Extra Tree Classifier), решающие деревья (Decision Tree), логистическая регрессия (Logistic Regression), градиентный бустинг (Gradient Boosting), градиентный бустинг на деревьях (XGB), случайные леса (Random Forest), метод опорных векторов (SVM), нейронные сети (Neural Network).

При первом ансамблировании использовано 3 метода машинного обучения, а именно: случайные леса (Random Forest), метод опорных векторов (SVM), нейронные сети (Neural Network) [3]

Результат ансамблирования приведём в метрике AUC [2].

Ниже приведен код на языке Python, с использованием библиотек для анализа данных (Pandas, SkLearn), написанный для прогнозирования оттока клиентов.

```
clf4 = RandomForestClassifier()
clf5 = SVC()
clf6 = MLPClassifier()
eclf3 = VotingClassifier(estimators=[('lr', clf4), ('rf', clf5), ('gnb', clf6)])
eclf3.fit(X_train, y_train)
predictions2 = eclf3.predict(X_test)
print(accuracy_score(y_test, predictions2))
```

По результатам обучения модель выдает верный результат в 79,9% случаях.

При втором ансамблировании использовано 3 метода машинного обучения, а именно: логистическая регрессия (Logistic Regression), градиентный бустинг (Gradient Boosting), градиентный бустинг на деревьях (XGB).

Результат ансамблирования приведём в метрике AUC.

Ниже приведен код на языке Python, с использованием библиотек для анализа данных (Pandas, SkLearn).

```

clf1 = XGBClassifier()
clf2 = LogisticRegression()
clf3 = AdaBoostClassifier()
eclf1 = VotingClassifier(estimators=[('lr', clf1), ('rf', clf2), ('gnb', clf3)],
voting='soft')
eclf1.fit(X_train, y_train)
predictions1 = eclf1.predict(X_test)
print(accuracy_score(y_test, predictions1))

```

По результатам обучения модель выдает верный результат в 81,37% случаях.

При третьем ансамблировании использовано 3 метода машинного обучения, а именно: наивный байесовский классификатор (Naive Bayes), случайные деревья (Extra Tree Classifier), решающие деревья (Decision Tree).

Результат ансамблирования приведём в метрике AUC.

Ниже приведен код на языке Python, с использованием библиотек для анализа данных (Pandas, SkLearn).

```

clf7 = GaussianNB()
clf8 = ExtraTreesClassifier()
clf9 = DecisionTreeClassifier()
eclf2 = VotingClassifier(estimators=[('lr', clf7), ('rf', clf8), ('gnb', clf9)],
voting="soft")
eclf2.fit(X_train, y_train)
predictions3 = eclf2.predict(X_test)
print(accuracy_score(y_test, predictions3))

```

По результатам обучения модель выдает верный результат в 76,78% случаях.

Таблица 1 - Сравнение ансамблей

	Первый метод	Второй метод	Третий метод	Результат
1	логистическая регрессия (Logistic Regression)	градиентный бустинг (Gradient Boosting)	градиентный бустинг на деревьях (XGB)	81,37%
2	случайные леса (Random Forest)	метод опорных векторов (SVM)	нейронные сети (Neural Network)	79,90%
3	наивный байесовский классификатор (Naive Bayes)	случайные деревья (Extra Tree Classifier)	решающие деревья (Decision Tree)	76,78%

В ходе выполнения научно-исследовательской работы осуществлён анализ ансамблей методов машинного обучения для прогнозирования оттока клиентов. Наилучший результат даёт результат ансамбль из 3 методов

машинного обучения: наивный байесовский классификатор (Naive Bayes), случайные деревья (Extra Tree Classifier), решающие деревья (Decision Tree). Данный ансамбль выдаёт верный результат в 81,37% случаев.

### Библиографический список

1. Dataset // Kaggle. URL: <https://www.kaggle.com/blastchar/telco-customer-churn> (дата обращения: 02.12.2019).
2. Кривая ошибок // Learn Machine Learning. URL: [http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D1%80%D0%B8%D0%B2%D0%B0%D1%8F\\_%D0%BE%D1%88%D0%B8%D0%B1%D0%BE%D0%BA](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D1%80%D0%B8%D0%B2%D0%B0%D1%8F_%D0%BE%D1%88%D0%B8%D0%B1%D0%BE%D0%BA) (дата обращения: 09.12.2019).
3. Ансамбли методов машинного обучения // Learn Machine Learning. URL: [https://learnmachinelearning.wikia.org/ru/wiki/%D0%90%D0%BD%D1%81%D0%B0%D0%BC%D0%B1%D0%BB%D0%B8\\_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D0%B5%D0%B9](https://learnmachinelearning.wikia.org/ru/wiki/%D0%90%D0%BD%D1%81%D0%B0%D0%BC%D0%B1%D0%BB%D0%B8_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D0%B5%D0%B9) (дата обращения: 16.12.2019).