|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство образования и науки Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К КУРСОВОЙ РАБОТЕ***

***НА ТЕМУ:***

***Решение комплексной задачи машинного обучения с учителем***

Студент \_\_ИУ-5 33М\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_А.П.Корнеева\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Преподователь **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_Ю.Е.Гапанюк**

(Подпись, дата) (И.О.Фамилия)

2021 г.

Содержание

[Введение 3](#_Toc90906938)

[Основная часть 4](#_Toc90906939)

[Пропуски 7](#_Toc90906940)

[Кодирование категориальных признаков 8](#_Toc90906941)

[Построение графика метрик качества модели 10](#_Toc90906942)

[AutoML 14](#_Toc90906943)

[Вывод 17](#_Toc90906944)

[Список литературы 17](#_Toc90906945)

# Введение

Аннотация

В ходе выполнения курсовой работы был проведён анализ выбранного набора данных, а также произведено предсказание размеров суммы трат покупателя. В работе содержится 17 страниц и 27 рисунков.

Актуальность

Результаты исследования являются востребованными и актуальными, т.к. компаниям, продающим свою продукцию, необходимо иметь результаты прогноза относительно размеров предполагаемой суммы трат покупателя, чтобы рассчитывать стоимость товаров.

Цель работы

Решить комплексную задачу машинного обучения с учителем.

Методы

Анализ и синтез.

# Основная часть

В ходе работы было произведено исследование набора данных, чтобы удалить имеющиеся пропуски для исключения искажения прогноза. Далее было выполнено кодирование категориальных признаков, а именно:

* «Gender»;
* «Partner»;
* «Dependents»;
* «PhoneService»;
* «MultipleLines»;
* «InternetService»;
* «OnlineSecurity»;
* «OnlineBackup»;
* «DeviceProtection»;
* «TechSupport»;
* «StreamingTV»;
* «StreamingMovies»;
* «Contract»;
* «PaperlessBilling»;
* «PaymentMethod»;
* «Churn».

Следующим шагом являлось обучение модели с различными вариантами. Было произведено построение графиков метрик качества модели, а также модели с использованием AutoML.

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

## Пропуски

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Пропусков в наборе данных не обнаружено.

## Кодирование категориальных признаков

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

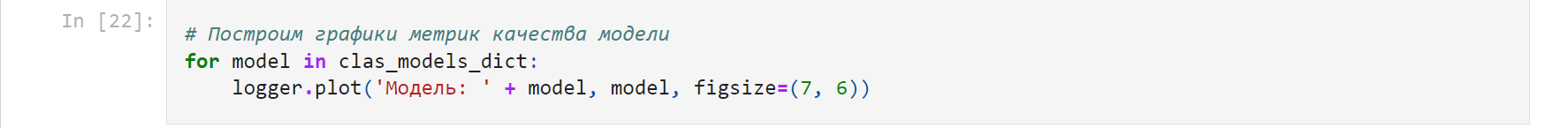
Изображение выглядит как текст

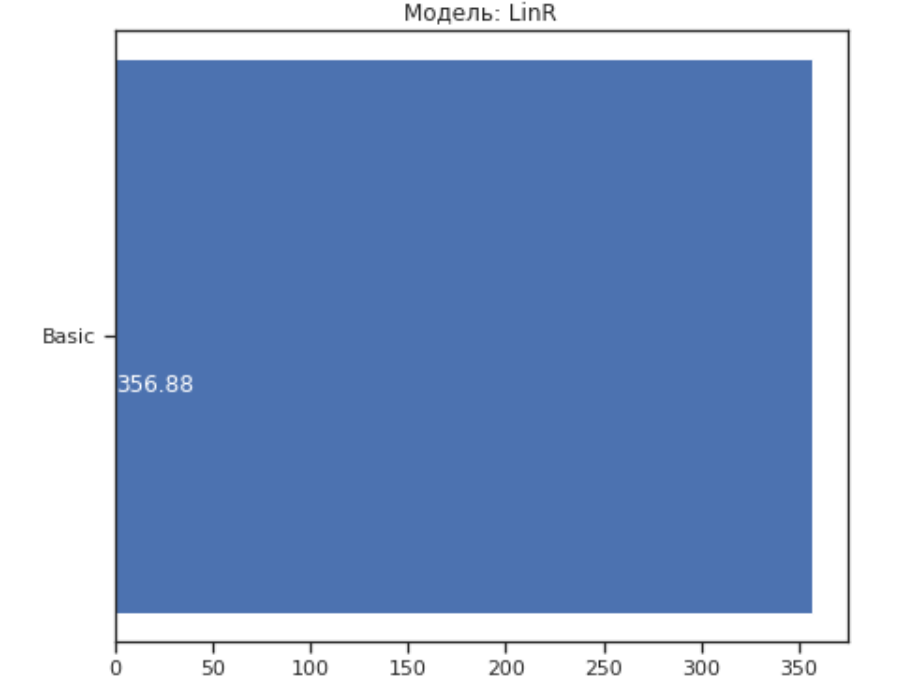
Автоматически созданное описание

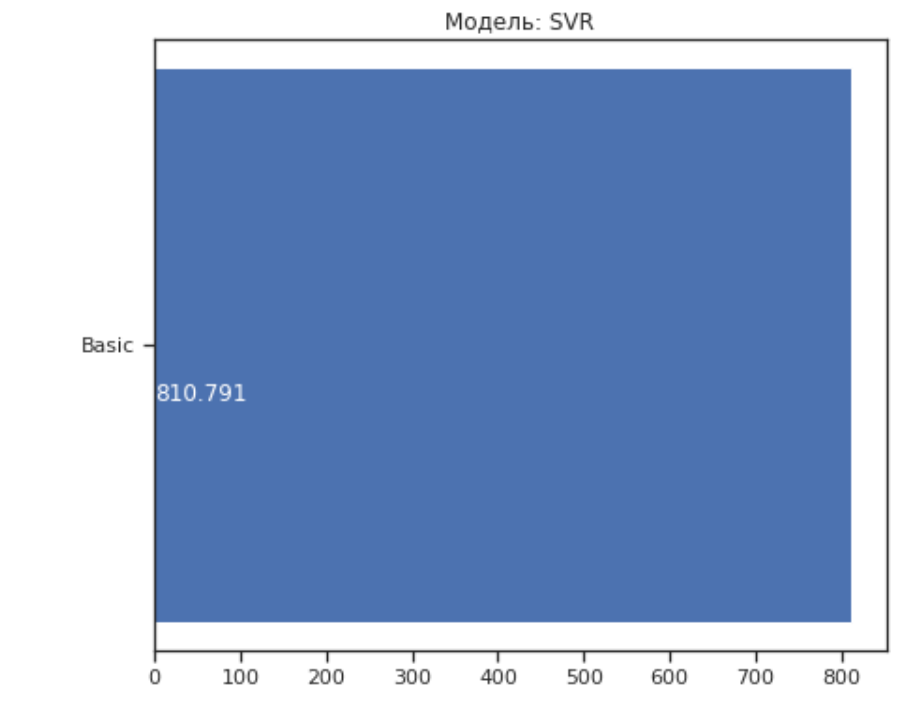
Изображение выглядит как текст

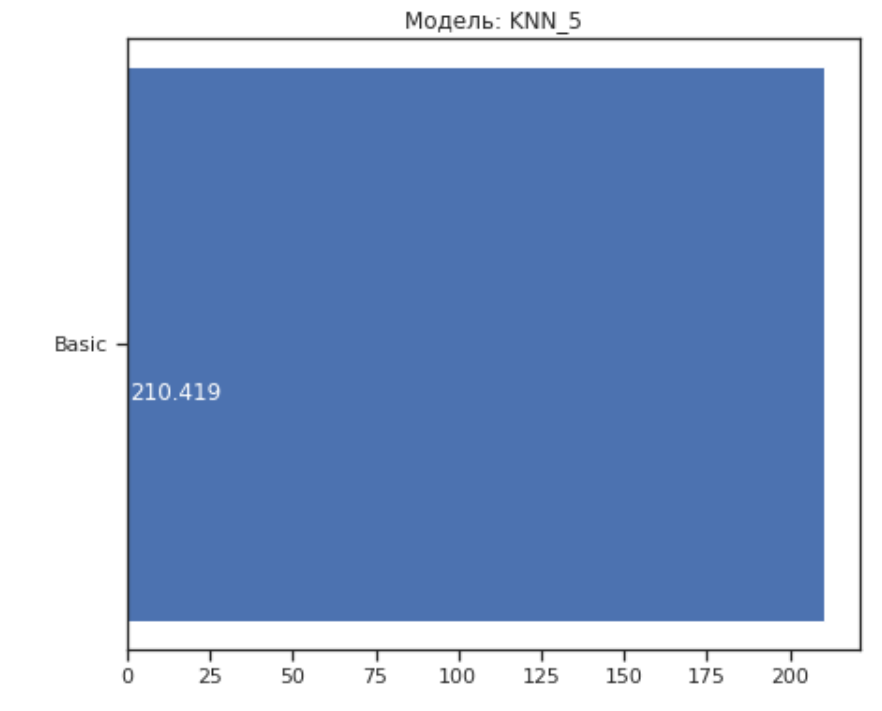
Автоматически созданное описание

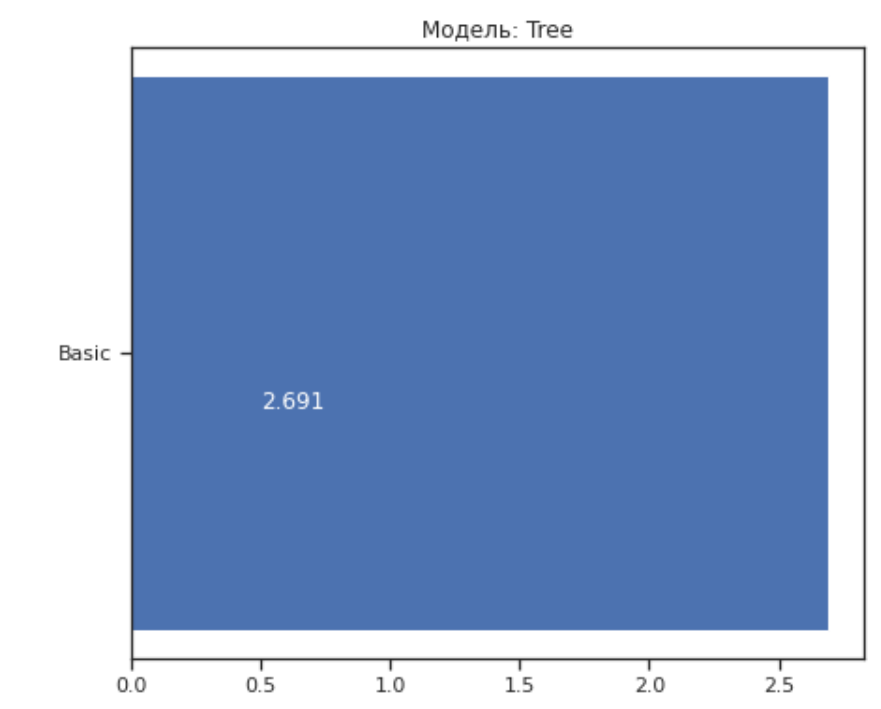
## Построение графика метрик качества модели

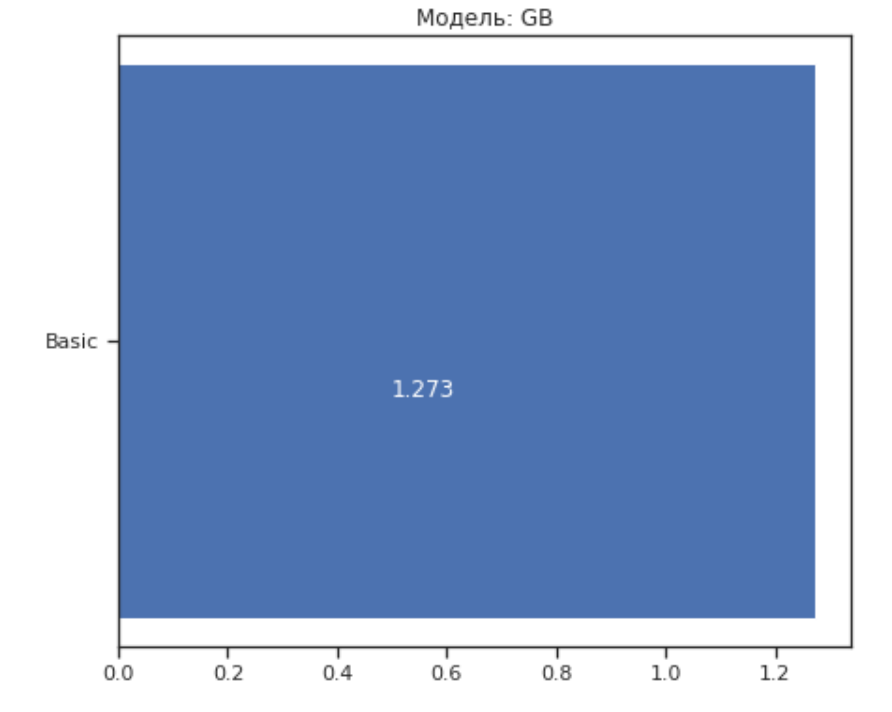


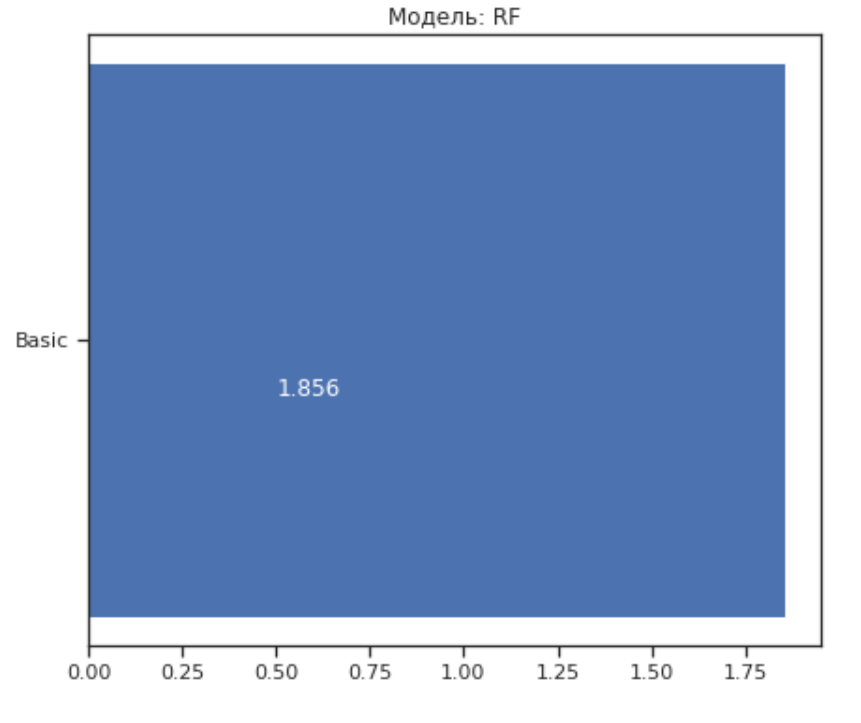












## AutoML

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как стол

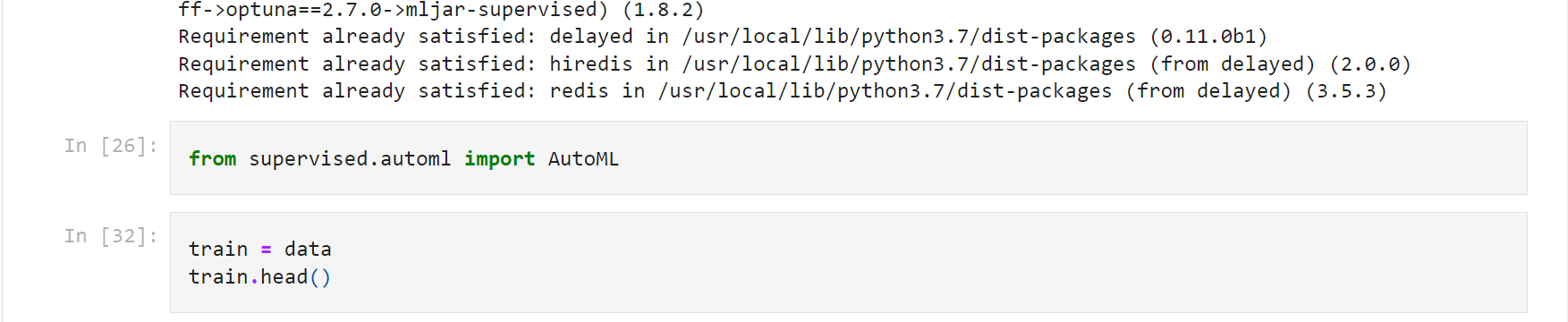
Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание



Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

# Вывод

В ходе выполнения курсовой работы была проведена проверка набора данных на предмет наличия пропусков, было произведено кодирование категориальных признаков, а также был построен график метрик качества модели и произведена работа с AutoML. Следовательно, цель работы достигнута.

# Список литературы

1. <https://www.automl.org/>
2. [Python: категориальные признаки | Анализ малых данных (dyakonov.org)](https://dyakonov.org/2016/08/03/python-%D0%BA%D0%B0%D1%82%D0%B5%D0%B3%D0%BE%D1%80%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5-%D0%BF%D1%80%D0%B8%D0%B7%D0%BD%D0%B0%D0%BA%D0%B8/)
3. [Метрики в задачах машинного обучения / Хабр (habr.com)](https://habr.com/ru/company/ods/blog/328372/)