**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**Национальный исследовательский ядерный университет «МИФИ»**

**ОТЧЕТ**

об исследовании

**Выполнил:** \_студент группы Б19-901\_\_ Бондаренко А.А.\_\_\_\_\_\_\_  
 (подпись) (Ф.И.О)

**Научный руководитель:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Смирнов Д.С.\_\_\_\_\_\_\_  
 (оценка) (подпись) (Ф.И.О)

Москва 2023 г

**Оглавление**

[1.Исходные данные, цели и задачи 2](#_Toc122983216)

[2. Выявление факторов, влияющих на уровень счастья. 2](#_Toc122983217)

[3. Обучение и тестирование моделей 6](#_Toc122983218)

[3.1. Линейная модель 6](#_Toc122983219)

[3.2. Полиномиальная модель 7](#_Toc122983220)

[3.3. Модель на основе случайного леса 8](#_Toc122983221)

4. Вывод

# **1.Исходные данные, цели и задачи**

Имеются данные с результатами опроса населения о его условиях существования. Переменные разбиты на 2 класса - "Признаки состояния" - это субъективная оценка населения своего бытия и "Признаки причины" - объектные количественные признаки оценивающие жизнедеятельность индивида и социума в котором он проживает. Также в данных присутствует интегральная характеристика удовлетворенности человека жизнью – «Ощущаемое счастье» – для ее описания используется шкала Кантрила.

**Цель работы**: спрогнозировать характеристику «Ощущаемое счастье» для жителей, у которых она отмечена как "Неизвестно".

**Задачи работы**:

* Определить, какие из признаков состояния наиболее сильно связаны с интегральной оценкой счастья респондента;
* Определить, как влияют признаки причины на наиболее важные признаки состояния;
* Пользуясь найденными закономерностями, спрогнозировать «Ощущаемое счастье» у респондентов.

# **2. Выявление факторов, влияющих на уровень счастья.**

Для начала работы выделим из файла с данными строчки с неизвестным параметром счастья.

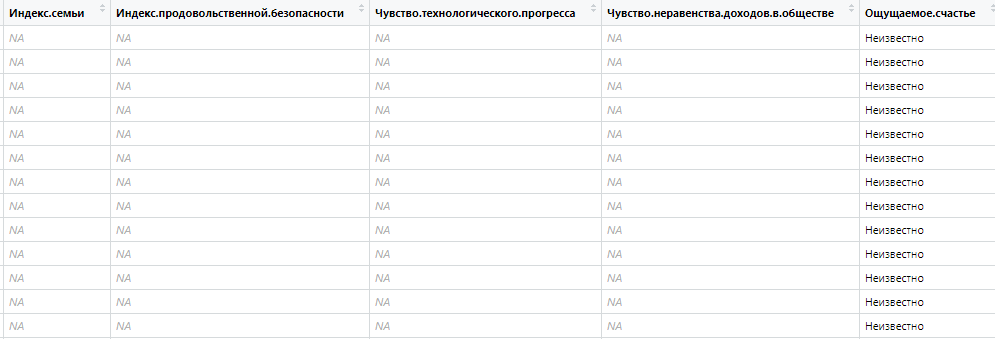


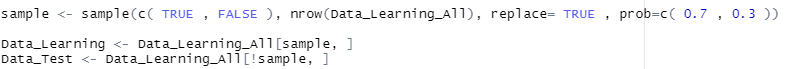
Таблица 2.1. Данные для предсказания (Data\_Forecast).

Конвертируем характеристику «Ощущаемое счастье» в численную шкалу от 1 до 10.



Таблица 2.2. Данные для обучения и тестов моделей (Data\_Learning\_All).

Разделим таблицу для предсказаний на две части в соотношении 70 на 30 соответственно для обучения моделей и тестирования точности прогнозов.



Чтобы определить, какие из признаков состояния наиболее сильно связаны с интегральной оценкой счастья жителя, была построена корреляционная матрица.

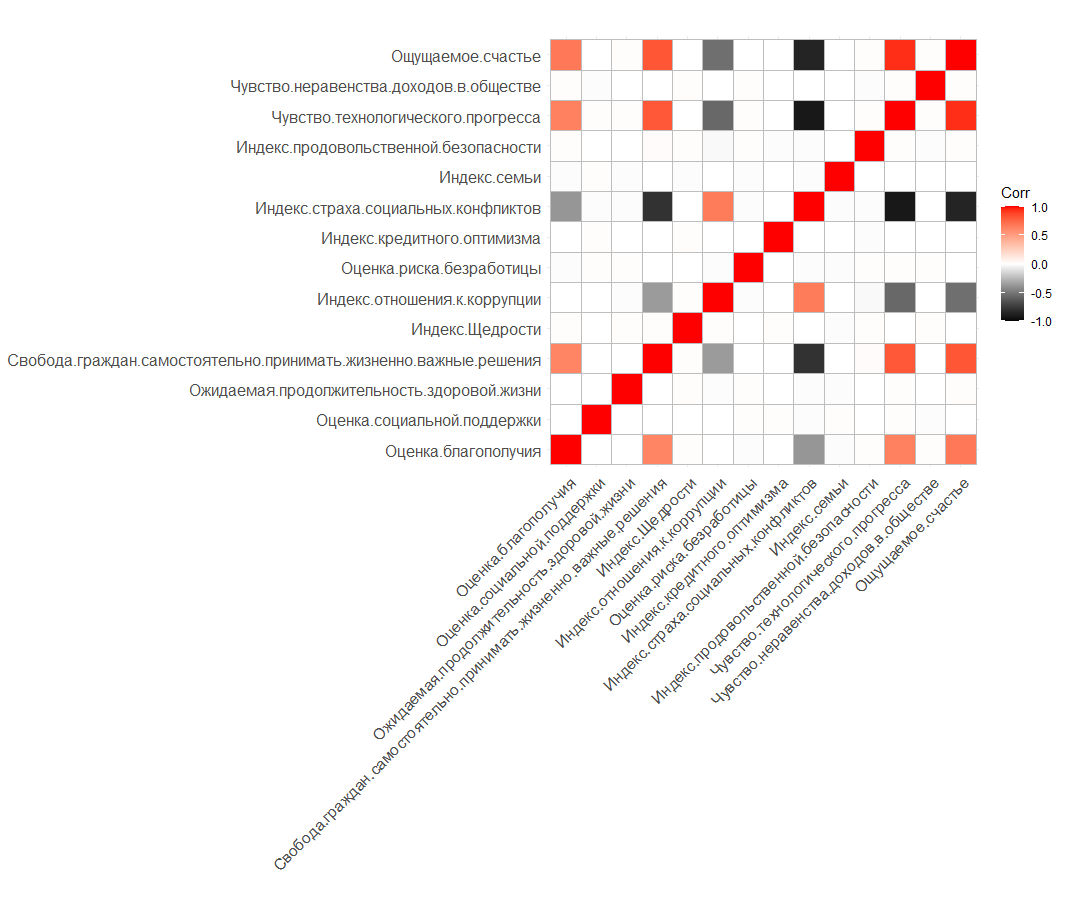


Таблица 2.3. Корреляционная матрица (Data\_Learning\_All).

Изображение выглядит как стол

Автоматически созданное описание



Таблица 2.4. Корреляционная матрица (Data\_Learning\_All).

Согласно корреляционной матрице, можно выделить 5 значимых признаков состояния:

Оценка благополучия= 0.68 (сильная связь)

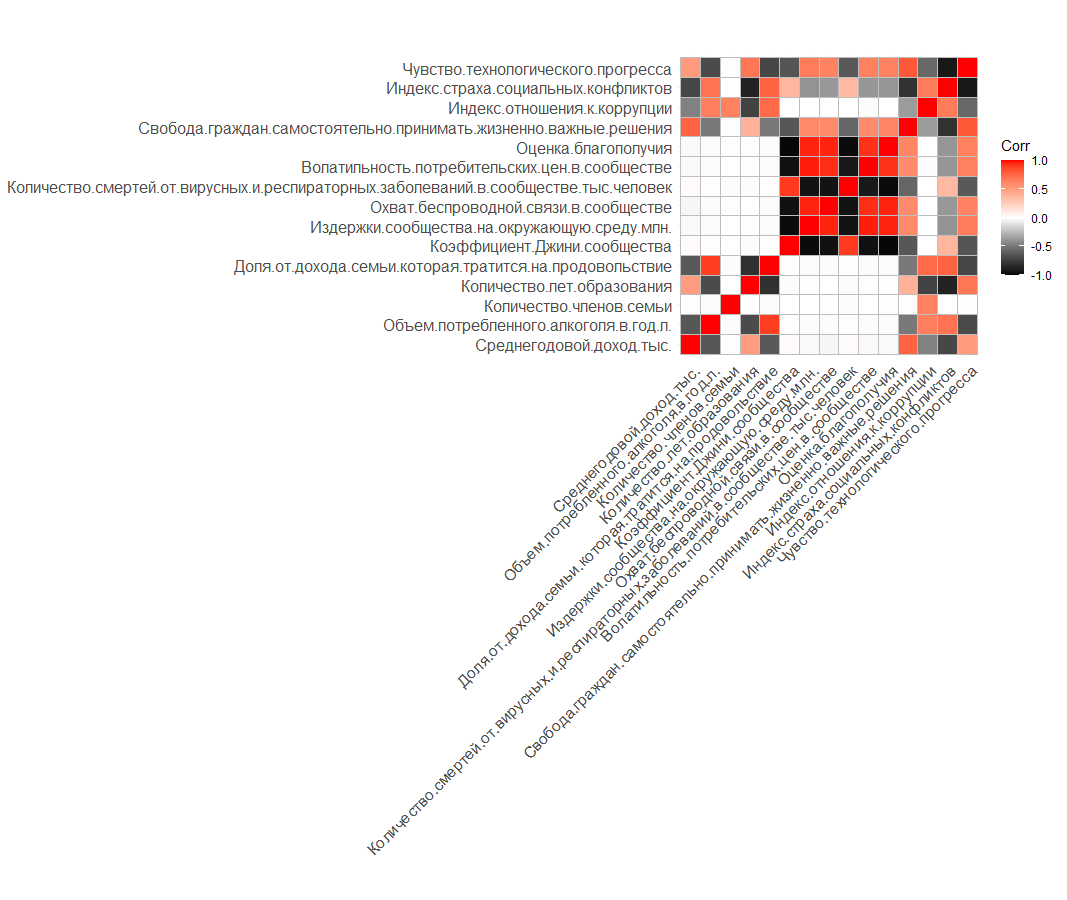
Свобода граждан принимать жизненно важные решения= 0.82 (сильная связь)

Индекс отношения к коррупции = -0.53 (умеренная связь)

Индекс страха социальных конфликтов= -0.86 (сильная связь)

Чувство технологического прогресса= 0.94 (сильная связь)

Далее были построены модели зависимостей каждого значимого признака состояния от признаков причины. При этом из всех признаков причины были взяты только самые значимые.



Оценка благополучия:

Изображение выглядит как текст

Автоматически созданное описание

Свобода граждан принимать жизненно важные решения:

Изображение выглядит как текст

Автоматически созданное описание

Индекс отношения к коррупции:

Изображение выглядит как текст

Автоматически созданное описание

Индекс страха социальных конфликтов:

Изображение выглядит как текст

Автоматически созданное описание

Чувство технологического прогресса:

Изображение выглядит как текст

Автоматически созданное описание

Для проверки моделей было осуществлено прогнозирование на тестовых данных. Точность результатов составила >99,7%

Поэтому было принято решение построения моделей на основе этих факторов.

# 

# **3. Обучение и тестирование моделей**

## **3.1. Линейная модель**

Была построена линейная модель зависимости «Ощущаемого счастья» от пяти признаков состояния.

Изображение выглядит как текст

Автоматически созданное описание

Модель дала точность прогноза в 44%.

## **3.2. Полиномиальная модель**

Затем была построена полиномиальная модель

Изображение выглядит как текст

Автоматически созданное описание

Путём многочисленных тестов удалось получить уровень точности в 49%.

## **3.3. Модель на основе случайного леса**

Затем в качестве эксперимента для улучшения точности прогноза была использована модель случайного леса, которая с первого раза дала точность в 89,3%.

Изображение выглядит как стол

Автоматически созданное описание

**4. Вывод.**

Определена итоговая модель по методу Случайного леса (Random Forest). Её тестирование проходила разбиением известных данных на 2 группы (70% для обучения и 30% для предсказания).

Итоговых респондентов (у которых счастье Неизвестно) модель предсказывает с точностью = 89.3%.

Ранее получена линейная модель, которая недостаточно хорошо (42 % точных попаданий) может предсказывать точный уровень Ощущаемого счастья у человека, основываясь на Индексе страха социальных конфликтов, Индексе отношения к коррупции, Оценке благополучия, Свободе граждан самостоятельно принимать жизненно важные решения, чувство технологического прогресса. Её тестирование проходило методом k-fold.

Выявлена закономерность, что чем выше Оценка благополучия, Свобода граждан самостоятельно принимать жизненно важные решения, чувство технологического прогресса, тем выше Ощущаемое счастье у людей. А Индекс страха социальных конфликтов, Индекс отношения к коррупции, наоборот, чем они выше, тем вероятнее, что у человека низкий уровень счастья.