

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дифференциальных уравнений и математической экономики

**ВИЗУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И РАЗРАБОТКА
СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ТЕКУЧЕСТИ
КАДРОВ**

БАКАЛАВРСКАЯ РАБОТА

Студента 4 курса 451 группы
Направления 38.03.05 – Бизнес-информатика
Механико-математического факультета
Ковина Семёна Дмитриевича

Саратов 2022

Цель бакалаврской работы: проведение визуального анализа данных о проблеме текучести кадров в АО «Неофлекс Консалтинг» и разработка системы прогнозирования вероятности ухода сотрудника из этой организации.

Задачи:

1. Изучить теоретическое обоснование проблематики текучести кадров.
2. Выполнить предобработку предоставленных данных о сотрудниках и подготовить их для машинного обучения.
3. Провести визуальный анализ данных и построить графики по статистическим показателям.
4. Построить модель классификации сотрудников относительно вероятности утечки кадра.
5. Построить регрессионную модель прогнозирования количества лет и месяцев, которые проработает сотрудник в данной организации.
6. Реализовать веб-приложение для удобного использования разработанных моделей прогнозирования.

Виды текучести кадров

1. Внутриорганизационная. Это движение персонала внутри одной компании, например, между отделами, департаментами или филиалами организации.
2. Внешняя. Передвижение работников между разными организациями, отраслями и сферами экономики.
3. Естественная текучесть персонала является нормальной и стимулирует процесс обновления коллектива. Если уровень текучести не превышает 3-5 процентов, то нет необходимости предпринимать какие-либо меры.
4. Психологическая, или скрытая текучесть кадров часто не рассматривается как реальная проблема. Такая текучесть персонала появляется, когда сотрудники официально продолжают работать, но по факту абстрагируются от рабочих процессов.
5. Физическая текучесть кадров. Увольнение сотрудников по любым другим причинам.

Входные данные о сотрудниках

Дата	Сотрудник	БН	grade
01.01.2019	8a195bab54f69a76d6be02a70692dc8e	ФАС	-1
01.01.2019	60af5ad43f95c071e5f3f3588281834c	Руководитель	9
01.01.2019	2e32f3195c2fd2d6660a4e48f5dbd40a	SOA-ВТБ	3
01.01.2019	c338b675efc7c06ca9f54c03c312ead1	BigData Solutions	2
01.01.2019	6f216663ad0259436e16b60c5202ce86	ФАС	-1
01.01.2019	68915ee9dd402ea9ab8745d3f1f1eca9	ФАС	-1
01.01.2019	b6c7af6f64f2e92c160caa1b4425247a	Водитель	-1
01.01.2019	1f3e490038d8dc794aaefa8802b34c0a	NFO	3
01.01.2019	bad50bad803a978392acd9d6922b7877	Департамент развития бизнеса	6
01.01.2019	b7eac6d9b1c222f3579036123b2b2256	SOA-ВТБ	2

Рисунок 1 – Фрагмент таблицы «Грейды»

Табельный номер	Сотрудник	Дата рождения	Дата приема	Дата увольнения	Физическое лицо.Пол
	06c8118e2c3726b66efb4b8d113e2635	15.03.1976	21.05.2009	30.06.2009	Мужской
6	06c8118e2c3726b66efb4b8d113e2635	15.03.1976	01.07.2009	28.06.2019	Мужской
	06c8118e2c3726b66efb4b8d113e2635	15.03.1976	01.07.2009		Мужской
7	4e24fb1750996e0e8f91861df9ff65e7	01.05.1979	02.07.2009		Женский
	4e24fb1750996e0e8f91861df9ff65e7	01.05.1979	02.07.2009		Женский
	80af072e4e86c60e06fede7a8de6fb79	11.04.1969	14.07.2009	14.09.2012	Мужской
11	23ec45eb079c9ef3fa719b144067ac66	01.06.1960	14.07.2009	29.11.2019	Мужской
	23ec45eb079c9ef3fa719b144067ac66	01.06.1960	14.07.2009		Мужской
	5e7b2efe12281cee893cb63580a64c77	20.10.1975	17.07.2009	07.04.2011	Мужской
	b97a0a84585008373abe3b888e86b359	09.12.1983	21.07.2009	25.10.2013	Мужской
	2d77f813f4db00eef993aca0642a13f0	05.05.1984	21.07.2009	20.03.2014	Мужской
	198643fbcec1dadd44bba2d09d80df71	02.05.1983	21.07.2009	27.02.2015	Мужской
	2ee0b7faeafbd6c0eac0577b7d819469	19.11.1984	21.07.2009	29.05.2015	Женский

Рисунок 3 – Фрагмент таблицы «Сотрудники»

Дата	Сотрудник	Филиал	Город
08.02.2005	60af5ad43f95c071e5f3f3588281834c	Москва	Москва
28.02.2005	e1e061c61daee1d518a9d8181e64c196	Москва	Москва
28.02.2005	8dcd51df55acd1afc191d263848e73d1	Москва	Москва
01.03.2005	16e30d6f664c5e4085b1e41e77b5aac0	Москва	Москва
02.08.2005	f66dac3272ab7d2398c79cbf9da03363	Москва	Москва
16.01.2006	7bd61b9428039f706f7c1e0602a5d5b1	Москва	Москва
07.12.2006	b2a67e7335e3072c1c6a01d5e7004c72	Москва	Москва
11.12.2006	82ac8834435110f81bb4e2dc2b12b567	Москва	Москва
02.04.2007	bead7915fc661af66f0862fff75eaa56	Москва	Москва

Рисунок 2 – Фрагмент таблицы «Города»

Подготовленный набор данных

	Дата приема	Дата увольнения	Ушел	Стаж	БН	Грейд	Филиал	Возраст	Пол
0	2013-07-10	2017-12-20	1	4.446361	Общее_Менеджеры	2.0	Москва	29	0
1	2013-07-10	2017-12-20	1	4.446361	Общее_Декрет	2.0	Москва	29	0
2	2020-11-24	NaT	0	0.476712	SOA	3.0	Воронеж	24	1
3	2017-03-01	NaT	0	4.213699	NFO	3.0	Саратов	29	0
4	2017-03-01	NaT	0	4.213699	Общее_Декрет	2.0	Саратов	29	0
...
2118	2013-06-10	2018-03-02	1	4.725627	Neoflex Reporting	2.0	Москва	27	0
2119	2013-06-10	2018-03-02	1	4.725627	Общее_Декрет	2.0	Москва	27	0
2120	2013-06-10	2018-03-02	1	4.725627	Neoflex Reporting ЦБ	2.0	Москва	27	0
2121	2013-06-10	2018-03-02	1	4.725627	Департамент аналитики	2.0	Москва	27	0
2122	2019-10-01	2020-11-27	1	1.158135	NFO	2.0	Саратов	19	1

Графическое изображение основных показателей

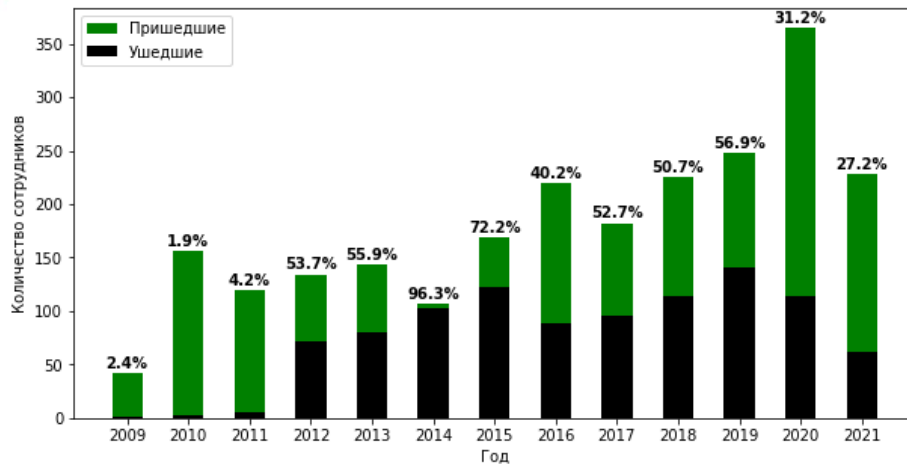


График 1 – Текучесть кадров по годам

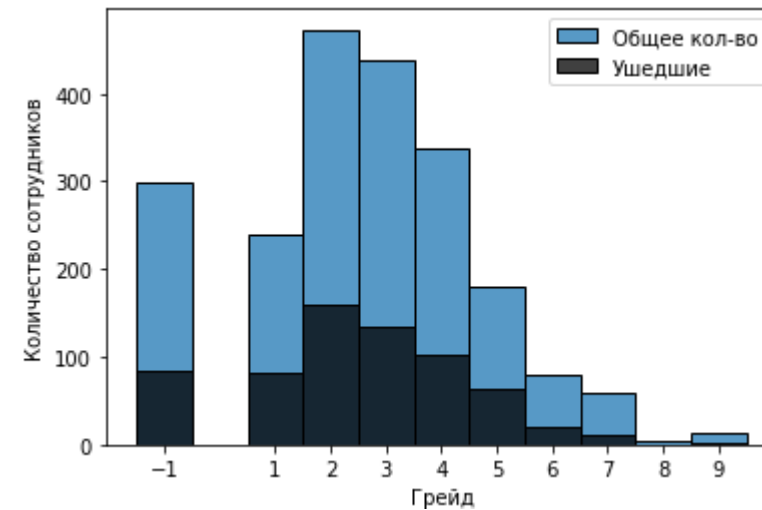


График 2 – Распределение сотрудников по грейдам

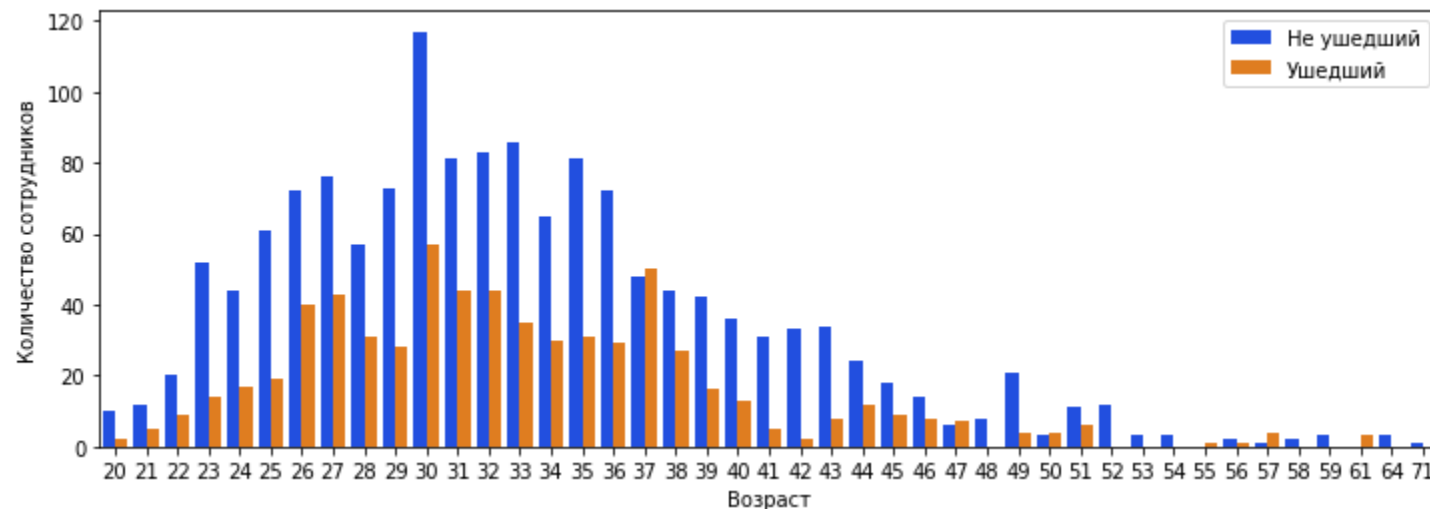
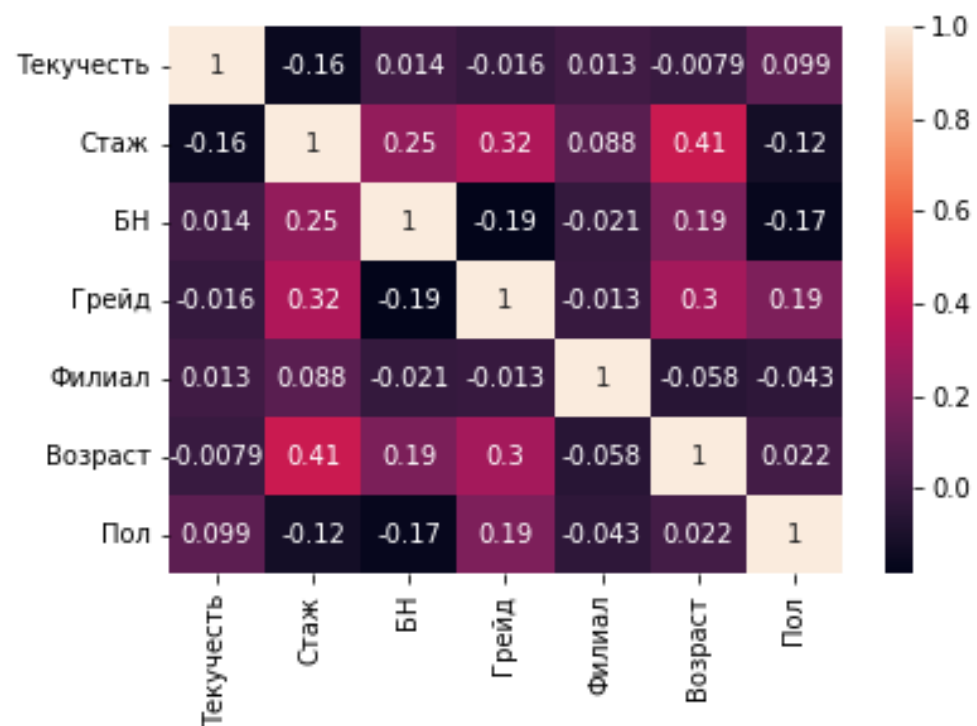


График 3 – Распределение ушедших и оставшихся сотрудников по возрасту

Отобранные признаки для предиктивных моделей и их корреляции



Подбор алгоритмов для обучения моделей

Модель 1. Классификация

	Algorithm	ROC AUC Mean	F1 Mean	Accuracy Mean
1	Random Forest	76.00	45.61	72.74
3	KNN	65.10	40.06	67.09
2	SVM	64.93	28.58	68.91
4	Decision Tree Classifier	63.18	49.11	67.96
5	Gaussian NB	60.28	6.47	68.22
0	Logistic Regression	60.09	45.91	55.15

Модель 2. Регрессия

	Algorithm	R2	MAE	MSE
1	Random Forest Regression	0.6081	1.56	4.80
2	Gradient Boosting Regression	0.5962	1.70	4.89
0	Linear Regression	0.2988	2.38	8.58
3	Ridge Regression	0.2988	2.38	8.58

Пример работы веб-сервиса (А)

Введите возраст сотрудника:

35 - +

Выберите пол сотрудника:

☐ Мужской

☒ Женский

В каком филиале работает сотрудник?

Саратов ▾

Бизнес-направление сотрудника:

Фин. рынки и риски ▾

Выберите грейд сотрудника:

6

-1 9

Введите количество лет, проработанных сотрудником:

4,5 - +

Рассчитать

Система прогнозирования текучести кадров

Для сотрудника с входными параметрами:

Возраст: 35

Пол: Женский

Город: Саратов

БН: Фин. рынки и риски

Грейд: 6

Стаж: 4.5 года

Вероятность ухода сотрудника: **13.0%**

Спрогнозированное количество лет, сколько сотрудник проработает: **4.5**

Сотрудник имеет **низкий** уровень риска текучести

Пример работы веб-сервиса (Б)

Введите возраст сотрудника:

20 - +

Выберите пол сотрудника:

☒ Мужской
☐ Женский

В каком филиале работает сотрудник?

Саратов ▾

Бизнес-направление сотрудника:

Фин. рынки и риски ▾

Выберите грейд сотрудника:

1

-1 9

Введите количество лет, проработанных сотрудником:

1,5 - +

Рассчитать

Система прогнозирования текучести кадров

Для сотрудника с входными параметрами:

Возраст: 20
Пол: Мужской
Город: Саратов
БН: Фин. рынки и риски
Грейд: 1
Стаж: 1.5 года

Вероятность ухода сотрудника: **48.0%**

Спрогнозированное количество лет, сколько сотрудник проработает: **0.8**

Сотурдник имеет **высокий** риск текучести

Спасибо за внимание!