

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

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

Кафедра дифференциальных уравнений и математи-  
ческой экономики

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

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

Студента 4 курса 451 группы  
направления 38.03.05 — Бизнес-информатика

механико-математического факультета

Ковина Семёна Дмитриевича

Научный руководитель

доцент, к. ф.-м. н.

\_\_\_\_\_

И. Ю. Выгодчикова

Заведующий кафедрой

профессор, д. ф.-м. н.

\_\_\_\_\_

С. И. Дудов

Саратов 2022

## ВВЕДЕНИЕ

Текучесть кадров - движение рабочей силы, обусловленное неудовлетворенностью работника рабочим местом или неудовлетворенностью организации конкретным работником. Сегодня текучесть кадров - одна из наиболее важных проблем, с которыми сталкиваются современные предприятия. Следует различать ее естественный уровень, который способствует обновлению производственных коллективов, происходит непрерывно и не требует каких-либо чрезвычайных мер со стороны кадровых служб и руководства, и повышенный, вызывающий значительные экономические потери.

В работе рассмотрено два класса задач: задача классификации сотрудника относительно вероятности утечки кадра и задача построения модели прогнозирования количества лет и месяцев, которое сотрудник проработает.

**Актуальность темы.** В современных условиях мало кто из руководителей российских предприятий задумывается о том, насколько дорого обходится предприятию текучесть кадров. Так, по результатам проведенного в 2018 году крупнейшим в мире сообществом HR-профессионалов – «Общество по управлению человеческими ресурсами» (SHRM) исследования, было выявлено, что стоимость замены сотрудников низшего уровня (в работы входит: поиск кандидатов, проведение с ними собеседований, прием на работу, обучение, работа над повышением производительности) составляет 30 – 50 % от их годового оклада. Замена сотрудников среднего звена обходится предприятиям еще дороже – до 150 % от их годового оклада, а топ менеджмента – до 400 %.

Таким образом, можно утверждать, что от уровня текучести персонала предприятия зависит, в том числе, его конечная прибыль, а также становление корпоративной культуры и успех бизнеса в целом. Безусловно, ликвидировать это явление в деятельности предприятия полностью невозможно, но его можно в какой-то степени контролировать и минимизировать.

**Целью бакалаврской работы** стали проведения визуального анализа данных о проблеме текучести кадров в АО «Неофлекс Консалтинг» и разработка системы прогнозирования вероятности ухода сотрудника из этой организации.

**Объектом исследования** является АО «Неофлекс Консалтинг».

**Предметом исследования** являются данные о сотрудниках и их текучести из организации АО «Неофлекс Консалтинг».

**Структура и содержание бакалаврской работы.** Работа состоит из введения, трех разделов, заключения, списка использованных источников и приложений. Перечень разделов ВКР:

1. Теоретические основы исследования текучести кадров;
2. Анализ данных о текучести кадров в компании АО «Неофлекс Консалтинг»;
3. Разработка системы прогнозирования текучести кадров.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**В первом разделе** рассмотрены теоретические основы изучения проблемы текучести кадров и современные подходы к их изучению. Даны определения понятий движение кадров, текучести кадров, штата и их математическая оценка.

Если рассматривать текучесть кадров как процесс обновления коллектива предприятия, то и с экономической точки зрения и с социально-организационной это непрерывный и естественный для организации и коллектива процесс.

Анализ текучести кадров подразумевает выявления аномальных показателей в ротации сотрудников и их градации. Далее выделены наиболее часто используемые показатели численности и состава сотрудников:

1. показатель среднесписочной численности сотрудников  $\underline{P} = \frac{0.5P_1 + P_2 + \dots + P_{11} + 0.5P_{12}}{12}$ , где  $P_1 \dots P_{12}$  – численность работников по месяцам;

2. коэффициент приема  $K_{\Pi} = \frac{P_{\Pi}}{\underline{P}}$ , где  $P_{\Pi}$  – численность принятых работников и  $\underline{P}$  – среднесписочная численность персонала;
3. коэффициент текучести кадров (выбытия)  $K_{В} = \frac{P_{УВ}}{\underline{P}}$ , где  $P_{УВ}$  – численность уволенных работников;
4. коэффициент стабильности  $K_{С} = \frac{1-P_{УВ}}{\underline{P}+P_{\Pi}}$ .

Разумеется, для каждой организации есть своя норма коэффициента текучести кадров, к достижению которой необходимо стремиться. Данная норма определяется на основании следующих факторов:

1. специфика отрасли;
2. категория персонала (линейный персонал или же управленческий);
3. конкурентоспособность организации;
4. территориальное расположение организации;
5. кадровая политика;
6. стиль управления.

Показатель интенсивности текучести кадров играет огромную роль при изучении особенностей текучести кадров в разрезе социальных, демографических, и профессиональных аспектов.

1. Коэффициент увольнений. Рассчитывается как отношение общего числа увольнение за отчетный период к среднесписочной численности персонала за этот период.
2. Индекс стабильности. Рассчитывается как отношение численности сотрудников, проработавших не менее одного года, к количеству сотрудников, принятых год назад.
3. Дополнительные индекс текучести. Рассчитывается как отношение численности сотрудников, поступивших на работу в течение одного года к средней численности персонала в течение одного года.

Выявлены основные причины текучести кадров и их пути решения, они изображены на рисунке 1.



Рисунок 1 - Причины и следствия текучести кадров

**Во втором разделе** представлен обзор и анализ данных о текучести кадров в компании АО Неофлекс.

Из исходных данных, предоставленных АО Неофлекс, в виде трех таблиц (таблица сотрудников, филиалов и бизнес направлений), был сформирован и очищен единый дата-фрейм для дальнейшего анализа и построения моделей машинного обучения. На рисунке 2 представлен

фрагмент дата-фрейма, полученного после проведения предварительной обработки 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

Рисунок 2 - Фрагмент подготовленного датафрейма

В качестве ключевых организационно-экономических характеристик можно выделить бизнес направления (БН), внутренний коэффициент уровня специалиста (Грейд) и город филиала (Филиал).

Из полученных данных следует, что есть 79 бизнес направлений, среди которых наибольшее количество сотрудников задействовано по направлениям:

- SOA;
- NFO;
- SOA-support;
- BigData Solutions;
- Neoflex Reporting.

Внутренний коэффициент уровня специалиста имеет значение -1 для интерна (т.е. еще нештатный сотрудник) и градацию от 1 до 9 для штатных сотрудников. Визуальное представление распределения коэффициентов уровня специалистов на основе исходного дата-фрейма, представлено ниже на рисунке 3.

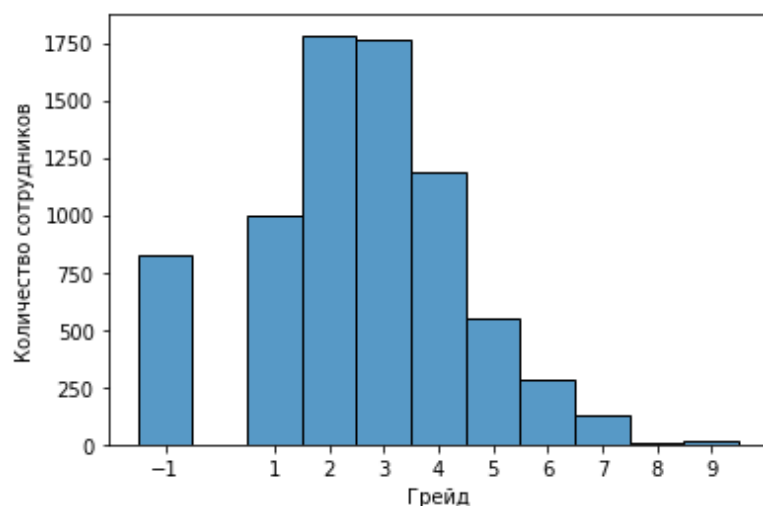


Рисунок 3 - Распределение грейдов сотрудников

Данные о филиалах предоставляют информацию о том в каких городах есть офисы АО Неофлекс, а именно:

- Саратов,
- Москва,
- Воронеж,
- Йоханнесбург,
- Пенза.

Из данных видно распределение по филиалам-городам, где на первом месте по количеству сотрудников оказалась Москва.

Также по подготовленным данным, после очистки и устранения проблемы дублирования сотрудников по причине небольших организационно-структурных нюансов, можно визуализировать распределение возраста ушедших и не ушедших из организации сотрудников, результат изображен на рисунке 4.

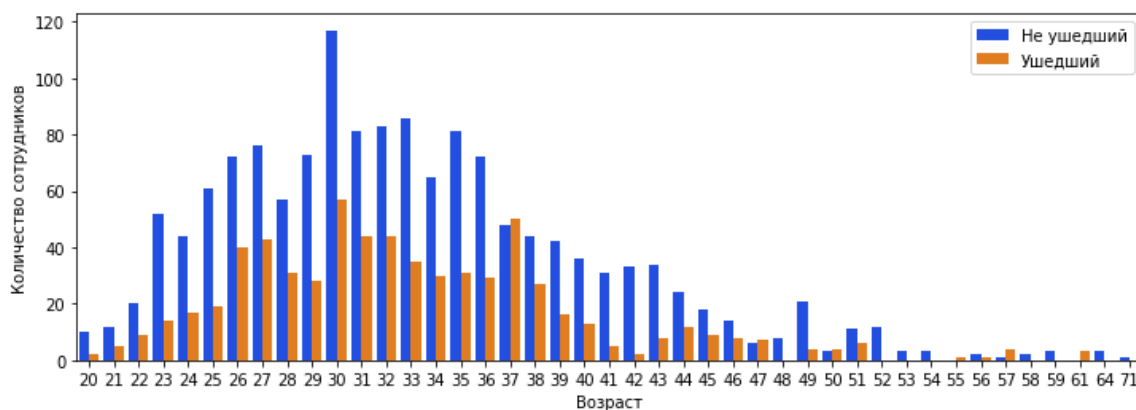


Рисунок 4 - Распределения возраста ушедших и не ушедших сотрудников

По рисунку выше можно выявить, что наибольшее количество сотрудников имеют возраст равному 30 годам. Причем в эту группу входит наибольшее количество как ушедших сотрудников, так и тех, кто продолжает свою деятельность в рассматриваемой компании.

Далее текучесть кадров была рассмотрена по каждому году деятельности компании АО «Неофлекс Консалтинг» в периоде с 2009 по 2021. Для построения визуализации считалось соотношение количества пришедших и ушедших сотрудников из компании за каждый год периода. График представлен в соответствии с рисунком 5.

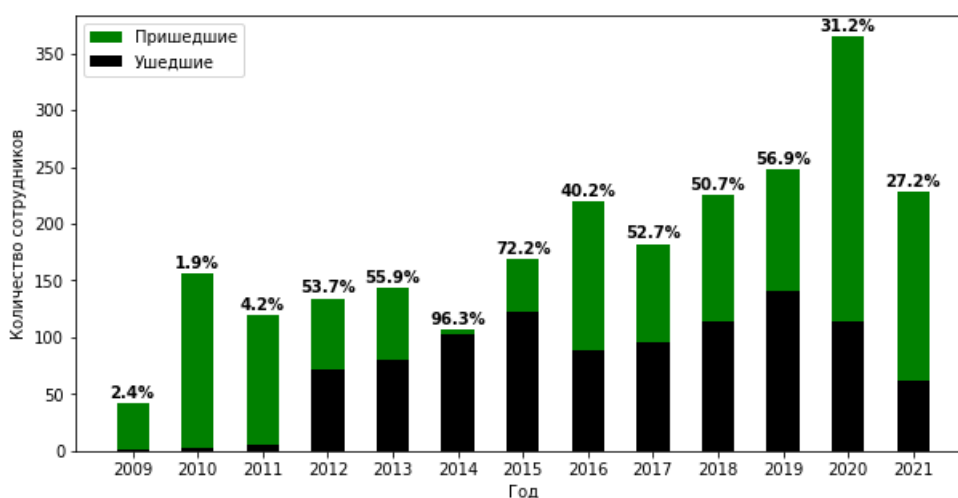


Рисунок 5 – Распределение текучести кадров по годам

На графике выше значение в 100% означает, что за n год из компании ушло столько же сотрудников, сколько пришло новых. По рисунку 5 явно видно, что в 2014 году компания потеряла сотрудников, практически равному числу пришедших (96.3%) за тот год.

Таким образом был выполнен этап предварительный обработки данных на предмет выбросов, дубликатов, аномалий и ошибок в исходных данных. Более того, проведен разведочный анализ данных по проблеме текучести кадров в компании АО «Неофлекс Консалтинг».

**Третий раздел** посвящен реализации, сравнению и выводу результатов работы моделей прогнозирования и классификации на основе данных из второго раздела.



В рамках работы были реализованы две модели. Первая представляет собой модель классификации сотрудника относительно вероятности утечки кадра или другими словами, расчет вероятности ухода сотрудника. Целью модели является общая оценка текучести кадров в организации и выявление наиболее вероятных на смену работы сотрудников.

Для выбора наиболее подходящего алгоритма машинного обучения, был проведен сравнительный анализ различных методов, как представлено на рисунке 6.

	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

Рисунок 6 - Сравнение работы алгоритмов машинного обучения

По проведенному сравнительному анализу различных методов машинного обучения, понятно, что наилучшим алгоритмом для обучения на подготовленных данных является «Случайный лес».

Вторая модель основана на задаче построения регрессии по предоставленным данным о рабочем персонале. Цель данной модели предсказать количество лет и месяцев, которые предположительно проработает новый сотрудник в компании, исходя из его пола, возраста, филиала, бизнес направления и внутреннего коэффициента уровня специалиста. Стоит отметить, что данная модель может быть применена также и к сотрудникам, которые уже проработали какое-то время в компании и может быть использована как дополнительный источник данных для первой модели.

В качестве использованной модели машинного обучения был выбран «Random Forest Regressor», так как при тестировании эта модель получила самую высокую оценку.

Для удобного взаимодействия с моделями было разработано веб-приложение с интерфейсом для ввода параметров сотрудника и выводом результатов оценки реализованной модели. В качестве основного инструмента реализации веб-приложения, был использован фреймворк Streamlit на языке Python.

Графики внутри веб-приложения реализованы при помощи библиотек Seaborn и Matplotlib на Python.

В результате при переходе по ссылке веб-приложения, загружается окно с формой для ввода данных и после нажатия кнопки «Рассчитать» введенные данные передаются модели, после чего выводится результат. Примеры работы системы прогнозирования представлены ниже на рисунках 7 и 8, причем на первом примере введены данные молодого специалиста.

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**

Сотрудник имеет **высокий** риск текучести

Рисунок 7 - Пример работы системы (А)

По рисунку выше видно, что сотрудник с низкими значениями возраста, грейда и стажа имеет высокий уровень риска ухода из компании.

На рисунке 8 можно увидеть, как работает прогноз системы на сотруднике с, наоборот, высокими значениями параметров.

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:

-1 6 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

Рисунок 8 - Пример работы системы (Б)

Так как у данного сотрудника значения грейда, возраста и стажа выше, то соответственно ниже вероятность его ухода из компании и проработает он дольше чем сотрудник А.

## ЗАКЛЮЧЕНИЕ

Основные результаты проведенной работы:

1. Изучено теоретическое обоснование проблематики текучести кадров.
2. Рассмотрены реальные данные компании АО Neoflex с точки зрения проблематики текучести кадров.
3. Проведена предобработка предоставленных данных и их визуальный анализ по статистическим показателям.
4. Построена модель классификации сотрудников относительно вероятности утечки кадра.
5. Построена модель прогнозирования количества лет и месяцев, которые проработает сотрудника в данной организации.
6. Реализовано веб-приложение для удобного использования разработанных моделей прогнозирования.