ВВЕДЕНИЕ

“«" - В современном бизнесе управление персоналом становится неотъемлемой частью стратегии компании из-за постоянных изменений и жесткой конкуренции. Сфера занятости становится все более динамичной, и сотрудники часто переходят из одной сферы работы в другую в поисках новых вызовов и возможностей для личностного развития.

Следует учитывать, что актуальность этой темы связана с необходимостью компаниям сокращать потери, связанные с увольнениями, и при этом сохранять свою конкурентоспособность. Глубокий анализ увольнений может выявить повторяющиеся тенденции или общие факторы, побуждающие сотрудников уходить, что позволяет разработать эффективные стратегии удержания персонала и улучшения рабочих условий.

Успешное функционирование компаний зависит от способности привлекать и удерживать талантливых специалистов, и анализ увольнений помогает определить пути предотвращения потери ценных кадров.

Цель данного исследования заключается в проведении анализа увольнений в ведущих компаниях за 2022 год с использованием методов моделирования линейной регрессии и методики Ridge.

Для достижения поставленной цели перед исследованием стоит ряд задач:

* анализ предметной области «Причины увольнения»;
* анализ алгоритма линейной регрессии с применением методики Ridge;
* выбор набора данных, содержащего информацию о причинах увольнений;
* построение сценария и логики обработки данных на основе алгоритма линейной регрессии с применением методики Ridge;
* формирование качественных выводов.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1 Увольнение сотрудников

Исследование причин увольнения сотрудников в современных организациях обычно основывается на разнообразных теориях и подходах, предложенных в области управления персоналом, организационного поведения и социологии труда. Основываясь на этом, проведем обзор некоторых ключевых теорий и подходов, которые формируют понимание причин увольнения персонала:

1. Теория удовлетворенности работой и мотивации.

Одной из основных теорий, оказывающей влияние на понимание увольнений, является теория удовлетворенности работой. Согласно этой теории, причины увольнения могут быть связаны с недостаточной удовлетворенностью работой, отсутствием возможностей карьерного роста или несоответствием корпоративной культуры. [1.1]

1. Теория ожидания и справедливости в вознаграждении.

Теория ожидания предполагает, что уровень удовлетворенности работой зависит от ожидаемой награды за труд и результаты, получаемые от организации. Если сотрудник не получает соответствующего вознаграждения или оценивает его как несправедливое, это может стать фактором увольнения.

1. Теория организационного поведения и конфликтов.

Организационное поведение и анализ конфликтов также играют важную роль в понимании причин увольнения. Например, конфликты с коллегами или руководством, недостаточная коммуникация или неэффективное управление могут значительно влиять на уровень удовлетворенности сотрудников своей работой и рабочей обстановкой. Если эти проблемы остаются неразрешенными или недооцененными, это может привести к тому, что сотрудники принимают решение о том, что их текущая рабочая среда не соответствует их ожиданиям или потребностям. В результате, это может стать фактором, приводящим к принятию решения об увольнении и поиску других мест для работы, где такие проблемы будут минимизированы или отсутствовать.

1. Теория человеческих ресурсов и кадровый менеджмент.

Теория человеческих ресурсов подчеркивает значение человеческого капитала для успеха организации. Она утверждает, что удовлетворенные и мотивированные сотрудники способствуют эффективной работе компании. Следовательно, факторы, влияющие на мотивацию и удовлетворенность, могут быть причинами увольнения, если они не удовлетворяются.

1. Теория организационной культуры.

Культура организации также играет значительную роль в удержании сотрудников. Если ценности компании не соответствуют ценностям сотрудника или существует несоответствие в ожиданиях, это может привести к увольнению.

Рассмотрим социально-экономические, организационные, психологические и культурные факторы, определяющие динамику увольнений:

1. Социально-экономические аспекты.

Один из основных социально-экономических факторов, влияющих на увольнение, — это уровень заработной платы и социальные льготы. Недостаточная оплата труда, отсутствие премий или возможностей для финансового роста могут стать факторами, мотивирующими сотрудников искать новые возможности на рынке труда. Также, общая экономическая ситуация в стране или регионе может сыграть роль. В периоды экономического кризиса или нестабильности на рынке труда, сотрудники могут быть склонны к увольнению из-за беспокойства о будущем или поиска более стабильного места работы. [1.2]

1. Организационные аспекты.

Структура и культура организации, методы управления и возможности профессионального роста также существенно влияют на решение сотрудников об увольнении. Недостаточная поддержка руководства, отсутствие возможностей для развития и карьерного роста, а также несоответствие корпоративной культуры ожиданиям сотрудников могут стать факторами, приводящими к решению об уходе из компании.

1. Психологические аспекты.

Мотивация, удовлетворенность работой, конфликты и стресс также оказывают влияние на принятие решения об увольнении. Например, если сотрудник не видит смысла в своей работе, испытывает постоянное недовольство или ощущает длительный стресс из-за рабочей среды, это может стать основой для принятия решения об уходе. [1.3]

1. Культурные аспекты.

Культурные различия и ценности также могут оказать влияние на увольнение сотрудников. Несоответствие ценностей компании и сотрудника, конфликты на этнической или культурной основе, а также отличия в представлениях о работе и ее целях могут стать причиной разрыва рабочих отношений. [1.4]

Анализ этих факторов необходим для понимания динамики увольнений в компаниях. Понимание мотивации увольнения поможет организациям принять меры по удержанию сотрудников, созданию более комфортной рабочей среды и улучшению условий труда для предотвращения потерь квалифицированных кадров. Все перечисленные аспекты важны для комплексного подхода к управлению персоналом и минимизации увольнений в современных организациях.

Увольнения сотрудников оказывают значительное воздействие на компанию, ее производительность, бренд и даже на оставшийся персонал. Изучение последствий увольнений позволяет понять, как это влияет на работоспособность компании:

* недостаток квалифицированных сотрудников, особенно в ключевых позициях, может привести к снижению производительности, задержкам в проектах и ухудшению качества работы;
* частые увольнения могут восприниматься как нестабильность в организации, что может отразиться на привлекательности компании для потенциальных сотрудников и клиентов;
* увольнения могут привести к снижению мотивации, ухудшению рабочего климата и даже к росту текучести кадров из-за неуверенности в будущем компании;
* компании, которые часто меняют состав своего персонала, могут восприниматься как место работы с нестабильной средой и ограниченными перспективами для карьерного роста.

Исследования показывают, что увольнения оказывают существенное влияние на долгосрочные перспективы компании. Они могут стать причиной финансовых потерь из-за потребности в новом найме, обучении новых сотрудников, а также из-за потери опыта и знаний, уходящих с уволенными сотрудниками.

Эффективное удержание персонала становится ключевой задачей для компаний, стремящихся сохранить высококвалифицированных и мотивированных сотрудников. Для достижения этой цели применяются различные стратегии, включающие программы мотивации, системы вознаграждения, развитие карьеры и управление производительностью:

1. Программы мотивации.

Компании часто используют разнообразные программы мотивации, чтобы стимулировать сотрудников к более эффективной работе и удерживать их в организации. Это может включать в себя предоставление бонусов за достижение целей, установление программ признания и поощрения, а также создание условий для личностного и профессионального роста.

1. Системы вознаграждения.

Эффективные системы вознаграждения играют ключевую роль в удержании сотрудников. Это включает уровень заработной платы, бонусные программы, акции компании, гибкие системы оплаты, предоставление льгот и привилегий, что способствует повышению удовлетворенности сотрудников и их привязанности к компании.

1. Методы развития карьеры.

Предложение сотрудникам перспектив развития внутри компании является мощным инструментом удержания.

1. Управление производительностью. [1.5]

Эффективные системы управления производительностью помогают установить четкие цели, оценивать и вознаграждать результаты, обеспечивая при этом контроль и обратную связь.

Индивидуальный подход и постоянное совершенствование стратегий управления персоналом представляют собой ключевые элементы для создания сильной и устойчивой команды. Однако, для достижения этой цели необходимо глубокое понимание методов анализа данных, их применение и последующее использование в исследовании увольнений.

1.2 Методы анализа данных и их применение в исследовании увольнений

Исследование увольнений сотрудников требует глубокого анализа данных, для чего ранее использовались различные методы, включая статистические подходы и машинное обучение на платформе Loginom. В данной работе мы переходим к использованию языка программирования R для реализации более продвинутого анализа с помощью линейной регрессии. Рассмотрим этапы исследования:

1. Статистические методы.

Прежде всего, будет использоваться дескриптивная статистика для анализа базовых характеристик увольнений, таких как среднее количество увольнений, их распределение по времени и отраслям. Это поможет установить фундамент для последующего анализа.

1. Машинное обучение.

Для более глубокого понимания динамики увольнений применяется модель линейной регрессии с использованием регуляризации Риджа в R. Это позволяет учесть возможную мультиколлинеарность между предикторами и повысить стабильность модели. Использование такого подхода позволяет не только выявить значимые факторы, влияющие на увольнения, но и улучшить качество предсказаний.

1. Инструменты визуализации.

В R доступен широкий спектр средств для визуализации данных, таких как ggplot2 для создания графиков и диаграмм, что обеспечивает наглядное представление результатов. Это позволит более эффективно представить результаты анализа и сделать выводы более доступными для понимания.

После того как данные собраны, их необходимо очистить и упорядочить, прежде чем приступать к их анализу.

Как только данные становятся структурированными и готовыми к анализу, особое внимание уделяется изучению временных рядов. Временной ряд (ряд динамики) — это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Анализ временных рядов позволяет выявлять закономерности и изменения в динамике показателя, что помогает в планировании, стратегических решениях и прогнозировании будущих значений.

В рамках исследования увольнений мы применяем линейную регрессионную модель для исследования зависимости между переменной отклика y и рядом предикторов x1,x2, …, xk-1. Модель представляет собой уравнение, где y выражается через линейную комбинацию предикторов плюс ошибка , которая учитывает неучтенные факторы и случайные флуктуации:

,

где b1, b2, …, bk-1 — параметры модели.

Для упрощения записи используется матричное представление:

,

где Y — вектор значений y;

X — матрица наблюдений;

B — вектор коэффициентов;

E — вектор ошибок.

В рамках анализа ставится задача оценки вектора B на основе данных. Обычно для этого используется метод наименьших квадратов, который минимизирует сумму квадратов разностей между наблюдаемыми и предсказанными значениями y. Если матрица XTX невырождена, то оценка вектора B вычисляется по формуле:

*.*

Часто приходится сталкиваться с ситуацией, когда матрица XTX почти вырожденная, что указывает на наличие мультиколлинеарности. В таких случаях оценки методом наименьших квадратов (МНК) могут быть формально верными, но обладать плохими статистическими свойствами. Даже небольшие изменения в исходных данных могут привести к значительным изменениям в оценках коэффициентов, включая изменение их знаков.

Метод регрессии Ридж включает регуляризацию для улучшения свойств оценок, позволяя преодолеть некоторые недостатки МНК. Когда зависимость между переменными почти линейна и количество предикторов значительно меньше объема выборки, МНК обычно дает хорошие результаты. Однако, если количество предикторов приближается к размеру выборки, увеличивается дисперсия прогнозов, что снижает их точность. Если количество предикторов превышает размер выборки, МНК не предоставляет уникального решения, делая дисперсию прогноза бесконечной. Регуляризация помогает уменьшить дисперсию прогноза за счет небольшого увеличения его смещенности, что улучшает общую точность прогноза.

Интерпретируемость модели также важна, особенно когда она содержит множество предикторов, многие из которых могут не влиять на отклик. Исключение незначительных переменных упрощает понимание модели. Ридж предлагает метод для уменьшения количества предикторов, уменьшая веса некоторых переменных до нуля, что делает модель более прозрачной. Ридж-регрессия определяет оценку неизвестного вектора B по стандартизованным наблюдениям с помощью уравнения:

,

где I — единичная матрица;

λ — параметр регуляризации.

Добавление λ к диагональным элементам матрицы XTX превращает плохо обусловленную матрицу в хорошо обусловленную, что позволяет избежать проблем с обращением плохо обусловленных матриц. Отличительной чертой Ридж-оценки является то, что она является МНК-оценкой с ограничением на норму возможных решений, что делает оценку смещенной, но более стабильной.

Таким образом, анализ в рамках нашего исследования позволяет глубже понять, как линейная регрессия Ридж может быть эффективно применена при проблемах мультиколлинеарности в данных. Добавление регуляризации через Ридж-регрессию улучшает свойства оценок, делая их более стабильными, даже если исходные данные показывают высокую степень коллинеарности между предикторами. Это достигается за счёт введения параметра λ, который контролирует величину штрафа за сложность модели, что помогает сбалансировать смещение и дисперсию оценок, предотвращая переобучение.

2 ПРАКТИЧЕСКАЯ ЧАСТЬ

2.1 Анализ коррелирующих факторов

В данной главе представлен анализ коррелирующих факторов в контексте исследования увольнений сотрудников за 2022 год. Целью данного исследования является оценка влияния различных факторов, таких как уровень удовлетворенности, последняя оценка, количество проектов, среднее количество часов в месяц и стаж работы в компании, на вероятность увольнения сотрудника.

Анализ коррелирующих факторов играет ключевую роль в понимании взаимосвязей между различными переменными и может помочь выявить наиболее важные факторы, влияющие на рассматриваемый результат. Для достижения этой цели использовались различные методы, включая расчет корреляционной матрицы, проведение статистических тестов корреляции и визуализацию данных с помощью диаграмм рассеяния.

С помощью функции cor() была рассчитана корреляционная матрица между переменными в данных, которые содержатся в переменной data. Эта функция предоставляется функционалом языка R и широко используется для анализа взаимосвязей между переменными. Для этого выбираются только числовые столбцы, соответствующие переменным "satisfaction\_level", "last\_evaluation", "number\_project", "average\_montly\_hours", "time\_spend\_company" и "salary".

Таблица 1 представляет собой таблицу корреляций, отображающую степень взаимосвязи между рассматриваемыми переменными: уровнем удовлетворенности сотрудников, последней оценкой, количеством проектов, средним количеством отработанных часов в месяц, стажем работы в компании и уровнем заработной платы. Для компактного представления все значения коэффициентов корреляции были округлены до трех знаков после запятой.

*Таблица 1 — Таблица корреляций*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Satisfaction**  **\_level** | **last\_**  **evaluation** | **number\_**  **project** | **average\_**  **montly\_**  **hours** | **time\_**  **spend\_**  **company** | **salary** |
| **satisfaction\_level** | 1 | 0.118 | –0.171 | –0.074 | –0.068 | 0.735 |
| **last\_**  **evaluation** | 0.118 | 1 | 0.548 | 0.528 | 0.366 | 0.444 |
| **number\_**  **project** | –0.171 | 0.549 | 1 | 0.636 | 0.379 | 0.131 |
| **average\_**  **montly\_**  **hours** | –0.074 | 0.528 | 0.636 | 1 | 0.359 | 0.393 |
| **time\_spend\_company** | –0.068 | 0.366 | 0.379 | 0.359 | 1 | 0.213 |
| **salary** | 0.735 | 0.444 | 0.131 | 0.393 | 0.213 | 1 |

Из таблицы корреляций видно, что наибольшая корреляция наблюдается между уровнем удовлетворенности сотрудников и их заработной платой (0.735). Это может указывать на то, что уровень заработной платы сильно влияет на уровень удовлетворенности сотрудников. Кроме того стоит отметить высокую степень корреляции между зарплатой и последней оценкой, что интуитивно понятно.

Когда коэффициент корреляции отрицательный, это указывает на то, что переменные двигаются в противоположных направлениях. Например, если уровень удовлетворенности сотрудников и количество проектов имеют отрицательную корреляцию, это означает, что более низкий уровень удовлетворенности может связываться с более высоким количеством проектов.

Наша же задача заключается в моделировании линейной регрессии уровня удовлетворенности сотрудника, поскольку этот показатель играет ключевую роль в принятии решения об увольнении. Поэтому важно изучить влияние различных факторов, таких как зарплата, последняя оценка и количество проектов, на уровень удовлетворенности сотрудников.

Для большей наглядности визуализируем корреляционные связи при помощи тепловой карты. Чем более насыщенный цвет в квадрате, тем сильнее корреляция между соответствующими переменными. Результат представлен на Рисунке 1.

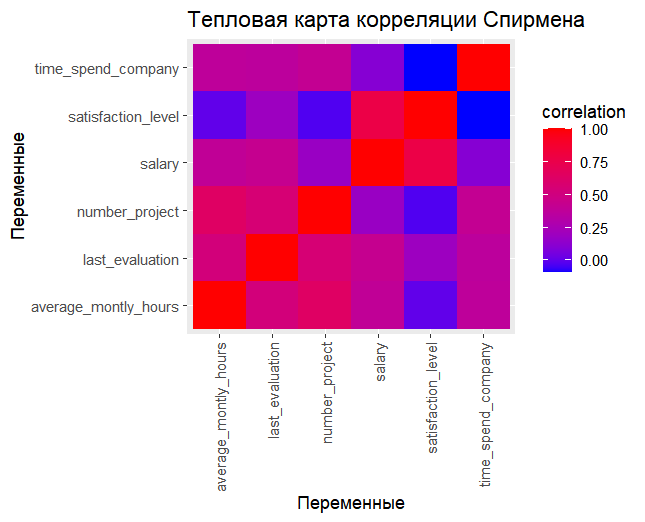


Рисунок 1 — Тепловая карта таблицы корреляций

На тепловой карте четко прослеживается высокая корреляция между уровнем заработной платы и уровнем удовлетворенности сотрудников, что подтверждает важность финансовых стимулов в формировании уровня удовлетворенности работников. Однако, менее очевидные корреляции, например между последней оценкой и уровнем удовлетворенности, также требуют внимания. В ходе проведенного теста корреляции Спирмена было выявлено, что р-value составил менее 0.05, что свидетельствует о статистически значимой корреляции между этими переменными. Это подтверждает наличие связи между последней оценкой сотрудника и его уровнем удовлетворенности.

На Рисунке 2 представлена диаграмма рассеяния, которая визуализирует эту корреляцию. На графике наблюдается кластерная структура и отчетливая тенденция увеличения уровня удовлетворенности с ростом последней оценки. Это дополнительно подтверждает наличие связи между этими двумя переменными и их влияние на удовлетворенность сотрудников.

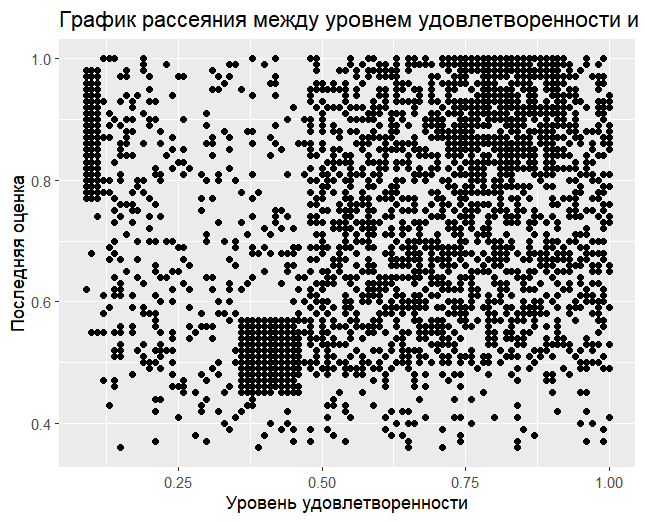


Рисунок 2 — График рассеяния между уровнем удовлетворенности и последней оценкой

В данной главе проанализированы взаимосвязи между различными переменными, влияющими на вероятность увольнения сотрудника. Используя методы корреляционного анализа, выявлены наиболее существенные факторы, такие как уровень удовлетворенности, зарплата и количество проектов. Эти результаты подтвердили важность финансовых стимулов и рабочей загруженности для уровня удовлетворенности сотрудников.

На следующем этапе продемонстрировано непосредственно моделирование линейной регрессии для более глубокого понимания влияния этих факторов на уровень удовлетворенности сотрудников и, соответственно, на вероятность их увольнения. С помощью регрессионного анализа появится возможность оценки степени влияния каждого фактора и предсказания уровня удовлетворенности сотрудника на основе имеющихся данных.

2.2 Построение модели линейной регрессии с применением методики Ridge

Для полного понимания методики Ridge в контексте линейной регрессии и ее применения к нашему исследованию, в данной главе будут разобраны шаги построения модели и оптимизации параметра регуляризации, а также проанализированы полученные результаты.

Первым этапом при построении линейной регрессии с применением метода регуляризации Ridge становится проверка на мультиколлинеарность, поскольку именно эта проблема может привести к нестабильности и неправильным выводам модели. Мультиколлинеарность — это состояние, когда независимые переменные в модели сильно коррелированы между собой.

Для этого была проведена проверка на мультиколлинеарность с помощью метода анализа дисперсии инфляции коэффициентов (VIF). Результаты проверки представлены в Таблице 2.

*Таблица 2 — Проверка на мультиколлинеарность*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **last\_evaluation** | **number\_project** | **average\_**  **montly\_hours** | **time\_spend\_**  **company** | **salary** |
| 1.847286 | 2.112113 | 2.053568 | 1.240625 | 1.432261 |

Мы можем наблюдать значения VIF для каждой из переменных: последней оценки, количества проектов, среднего количества отработанных часов в месяц, стажа работы в компании и уровня заработной платы. Низкие значения VIF, в основном около 1, свидетельствуют о невысокой степени мультиколлинеарности между независимыми переменными. Значения VIF порядка двух так же не свидетельствует о критичной корреляции, но все же применение регуляризации относительно коэффициентов модели поможет повысить ее устойчивость.

Для начала была построена модель линейной регрессии с применением Ridge, используя метод кросс-валидации для выбора оптимального значения параметра регуляризации. Регуляризация в модели Ridge добавляет штраф к коэффициентам регрессии, чтобы предотвратить переобучение модели. Параметр регуляризации (также известный как параметр λ или alpha) контролирует силу этого штрафа. Была применена кросс-валидация для того, чтобы найти такое значение параметра, которое обеспечивает наилучшее качество модели на независимых данных.

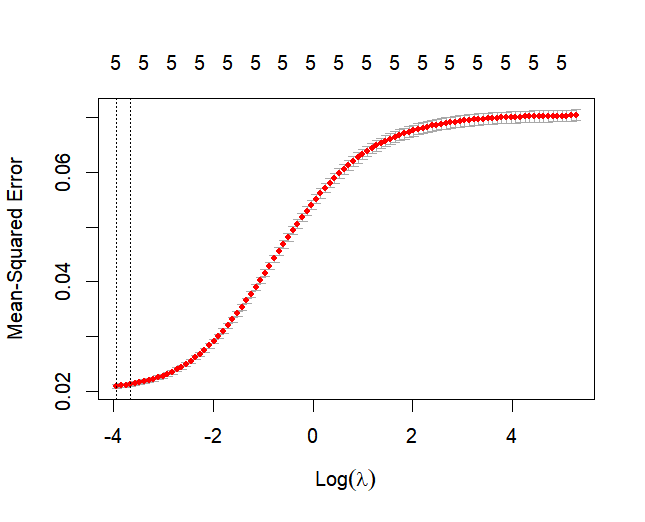
я

Рисунок 3 — Зависимость среднеквадратичной ошибки от значения параметра регуляризации

На Рисунке 3 представлена визуализация процесса выбора оптимального значения параметра регуляризации, засчет построения графика зависимости MSE от значения параметра. Этот график позволяет наглядно увидеть, как значение MSE меняется в зависимости от параметра регуляризации.

После построения модели регрессии мы анализируем полученное уравнение регрессии, чтобы понять, как каждый из предикторов влияет на целевую переменную. Наше уравнение регрессии выглядит следующим образом:

,

где ε — случайная ошибка.

Высокий положительный коэффициент при зарплате (0.288) отражает ранее обнаруженную корреляцию, свидетельствующую о том, что уровень удовлетворенности увеличивается с увеличением заработной платы. Однако, отрицательные коэффициенты, например, при number\_project, указывают на то, что увеличение этой переменной связано с уменьшением уровня удовлетворенности.

Высокое значение свободного члена указывает на высокий базовый уровень удовлетворенности сотрудников, даже если другие предикторы (переменные) имеют нулевое значение. Это может быть обусловлено, например, особенностями культуры компании, предоставляемыми льготами или условиями труда.

Далее мы оцениваем качество модели с помощью коэффициента детерминации (R-квадрат), который составляет 70 процентов. Это означает, что 70 % изменчивости уровня удовлетворенности сотрудников в наших данных может быть объяснено включенными в модель предикторами, такими как оценка работы, количество проектов, среднее количество часов в месяц, стаж работы и уровень заработной платы.

Дополнительно к этой оценке, важно отметить, что визуализация реальных и прогнозируемых значений уровня удовлетворенности сотрудников является ключевым инструментом для более глубокого анализа результатов модели. Несмотря на то, что не все наблюдения идеально совпадают с прогнозируемыми значениями из-за сложности факторов, влияющих на удовлетворенность сотрудников, мы видим, что общая тенденция моделирования схожа с реальными данными. Это говорит о том, что модель способна улавливать основные закономерности и тенденции, характеризующие уровень удовлетворенности сотрудников.

Таким образом, наши результаты подтверждают адекватность модели и ее способность к объяснению уровня удовлетворенности сотрудников в рамках данного исследования. Результат представлен на Рисунке 4.

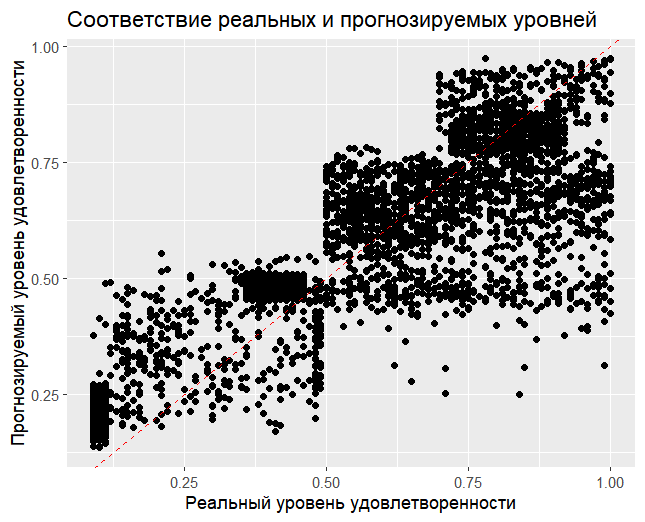


Рисунок 4 — График рассеяния реальных и прогнозируемых значений уровня удовлетворенности сотрудников

Исследование факторов, влияющих на уровень удовлетворенности сотрудников, имеет прямое отношение к анализу увольнений в компании. Это помогает выявить потенциальные причины, по которым сотрудники могут покидать организацию.

В данном контексте высокий уровень заработной платы, как показывает анализ, существенно влияет на уровень удовлетворенности сотрудников. Это подтверждает значимость финансового стимула для удержания персонала и снижения вероятности их увольнения. Кроме того, оценка других факторов позволяет выявить дополнительные аспекты, которые могут влиять на удовлетворенность сотрудников и, следовательно, на решение об увольнении. Например, низкая оценка или перегрузка проектами могут привести к демотивации сотрудников и увеличить вероятность их ухода из компании.

Таким образом, анализ уровня удовлетворенности сотрудников позволяет выявить ключевые факторы, способствующие увольнениям, и предоставляет базу для разработки стратегий по удержанию персонала и снижению текучести кадров.

ЗАКЛЮЧЕНИЕ

Анализ увольнений сотрудников представляет собой уникальный инструмент, который позволяет компаниям глубже понять внутренние динамики и тенденции, происходящие внутри организации. Значимость этого анализа заключается не только в том, чтобы предоставить текущую картину кадрового состава компании, но и в том, чтобы выявить широкий спектр причин, которые могут быть связаны с увольнениями. Благодаря этому компании имеют возможность не только реагировать на уход сотрудников, но и разрабатывать стратегии, направленные на укрепление и развитие кадрового потенциала. Путем изучения данных об увольнениях и причинах, лежащих в их основе, компании могут создать более эффективные методы управления персоналом, что способствует достижению устойчивого успеха и долгосрочного развития.

Цель — провести анализа увольнений в ведущих компаниях за 2022 год с использованием методов моделирования линейной регрессии и методики Ridge — достигнута.

В ходе выполнения данной курсовой работы были выполнены следующие задачи:

* проанализирована предметная область «Причины увольнения»;
* проанализирован алгоритм линейной регрессии с применением методики Ridge;
* выбран набор данных, содержащий информацию о причинах увольнений;
* построен сценарий и логика обработки данных на основе алгоритма линейной регрессии с применением методики Ridge;
* сформированы качественные выводы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

* 1. Веснин, В. P. Управление персоналом. Теория и практика: учебное пособие / В. P. Веснин. — Москва: ТК Велби, 2011. — 688 с. Дата обращения (05.05.2024).
  2. Иванцевич, Д.М. Человеческие ресурсы управления. / Д.М. Иванцевич, А.А. Лобанов — Москва: Дело, 2011. — 304 с. Дата обращения (05.05.2024).
  3. Кибанов, А.Я., Управление персоналом организации: Учебник/ Кибанов А.Я., Баткаева И.А., Ивановская Л.В.; Под ред. Кибанов А.Я., — 4-е изд., доп. и перераб. — М.:НИЦ ИНФРА-М, 2016. — 695 с. Дата обращения (05.05.2024).
  4. Маклаков, А.Г. Профессиональный психологический отбор персонала: Учебник для вузов. / А.Г.Маклаков. — СПб.: Питер, 2012. — 480 с. Дата обращения (09.05.2024).
  5. Пугачев, В.П. Руководство персоналом организации. / В.П. Пугачев — Москва: Аспект Прогресс, 2013. — 314 с. Дата обращения (15.05.2024).
  6. Семенов, В. Н. Экономика предприятия: Учебник / В.Н Семенов, 3-е изд., перераб. и доп. — М.: Центр экономики и маркетинга, 2015. — 360с. Дата обращения (15.05.2024).
  7. Шапиро С.А., Шатаева О.В. Основы управления персоналом в современных организациях: Экспресс-курс / С.А. Шапиро, О.В. Шатаева — Москва:ГроссМедиа, 2014. — 199 с. Дата обращения (16.05.2024).
  8. Бергер Е.Г. Нормоконтроль документации [Электронный ресурс]: Методические рекомендации / Бергер Е.Г., Зуев А.С. — М.: МИРЭА — Российский технологический университет, 2020. — 1 электрон. опт. диск (CD-ROM). Дата обращения (15.05.2024).
  9. Юрченко, Т. В. Эконометрика: временные ряды: учебное пособие / Т. В. Юрченко. — Санкт-Петербург: ИЭО СПбУТУиЭ, 2022. — 135 с. Яз. рус. Дата обращения (11.05.2024).
  10. Иванов, Б. Н. Теория вероятностей и математическая статистика: учебное пособие / Б. Н. Иванов. — 2-е изд., испр. и доп. — Санкт-Петербург: Лань, 2022. — 224 с. — ISBN 978-5-8114-3636-1. — Текст: электронный // Лань: электронно-библиотечная система. — Режим доступа: https://e.lanbook.com/book/206201 — Яз. рус. Дата обращения: (11.05.2024).

ПРАКТИЧЕСКАЯ ЧАСТЬ

1. Kaggle.com [Электронный ресурс] / Human Resource — <https://www.kaggle.com/datasets/kuniowu/human-resource> Дата обращения: (05.05.2024).
2. Официальный сайт R-Studio [Электронный ресурс] / R-Studio — Руководства R. — Режим доступа: https://cran.rstudio.com/manuals.html — Загл. с экрана. — Яз. рус. Дата обращения (06.05.2024).
3. Студенческий научный форум [Электронный ресурс] / Анализ технологий увольнения персонала — Режим доступа: https://scienceforum.ru/2015/article/2015010001. — Загл. с экрана. — Яз. рус. Дата обращения (10.05.2024).

ПРИЛОЖЕНИЯ

Приложение А — Программный код

Приложение А

В Листинге А.1 представлен итоговый программный код проекта на языке R.

*Листинг А.1 — Программный код линейной регрессии с применением методики Ridge*

|  |
| --- |
| #Установка необходимых пакетов  install.packages("readr")  install.packages("tidyr")  install.packages("vcd")  install.packages("glmnet")  install.packages("caret")  library(glmnet)  library(readr)  library(tidyr)  library(ggplot2)  library(vcd)  library(car)  library(caret)  library(corrplot)  #Загрузка датасета Human Resource  file\_path <- "HR.csv"  data <- read\_csv(file\_path,  col\_types = cols(  satisfaction\_level = col\_double(),  last\_evaluation = col\_double(),  number\_project = col\_integer(),  average\_montly\_hours = col\_integer(),  time\_spend\_company = col\_integer(),  salary = col\_integer()  ))  #Создание гистограмм для каждой числовой переменной  numeric\_columns <- sapply(data, is.numeric)  numeric\_data <- data[, numeric\_columns]  lapply(names(numeric\_data), function(x) {  p <- ggplot(data, aes\_string(x = x)) +  geom\_histogram(bins = 30, fill = "blue", color = "black") +  ggtitle(paste("Распределение переменной:", x))  print(p)  }) |

*Продолжение листинга А.1*

|  |
| --- |
| #Для применения корреляции Пирсона необходимо проверка на нормальность  numeric\_columns <- sapply(data, is.numeric)  results <- lapply(data[, numeric\_columns], shapiro.test)  print(results)  #Рассчет корреляции Спирмена  correlation\_matrix\_spearman <- cor(data[, c("satisfaction\_level", "last\_evaluation", "number\_project", "average\_montly\_hours", "time\_spend\_company", "salary")], method = "spearman")  #Преобразование корреляционной матрицы в датафрейм для ggplot  correlation\_df <- as.data.frame(correlation\_matrix\_spearman)  correlation\_df$variables <- rownames(correlation\_df)  correlation\_df\_long <- tidyr::gather(correlation\_df, key = "variable2", value = "correlation", -variables)  #Создание тепловой карты  ggplot(correlation\_df\_long, aes(x = variables, y = variable2, fill = correlation)) +  geom\_tile() +  scale\_fill\_gradient(low = "blue", high = "red") +  labs(title = "Тепловая карта корреляции Спирмена",  x = "Переменные", y = "Переменные") +  theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))  #Рассчет корреляции между уровнем удовлетворенности и последней оценкой  cor\_spearman <- cor.test(data$satisfaction\_level, data$last\_evaluation, method = "spearman")  print("Корреляция Спирмена:")  print(cor\_spearman)  #График рассеяния между уровнем удовлетворенности и последней оценкой  scatter\_plot <- ggplot(data, aes(x = satisfaction\_level, y = last\_evaluation)) +  geom\_point() +  labs(x = "Уровень удовлетворенности", y = "Последняя оценка") +  ggtitle("График рассеяния между уровнем удовлетворенности и последней оценкой")  print(scatter\_plot)  #Проверка на мультиколлинеарность  model <- lm(satisfaction\_level ~ last\_evaluation + number\_project + average\_montly\_hours + time\_spend\_company + salary, data=data)  vif(model) |

*Продолжение листинга А.1*

|  |
| --- |
| #Извлечение целевой переменной и признаков для модели  x <- as.matrix(data[, c("last\_evaluation", "number\_project", "average\_montly\_hours", "time\_spend\_company", "salary")])  y <- data$satisfaction\_level  model <- glmnet(x, y, alpha = 0 )  #Модель Ridge с помощью кросс-валидации для выбора оптимального значения параметра регуляризации  ridge\_model <- cv.glmnet(x, y, alpha = 0)  #Определение оптимального значения параметра регуляризации  lambda\_optimal <- ridge\_model$lambda.min  plot(ridge\_model)  lambda\_optimal  #Построение модели Ridge с оптимальным значением параметра  final\_ridge\_model <- glmnet(x, y, alpha = 0, lambda = lambda\_optimal)  #Вывод уравнения регрессии  coefficients <- as.vector(coef(final\_ridge\_model, s = lambda\_optimal)[-1])  intercept <- coef(final\_ridge\_model, s = lambda\_optimal)[1]  variable\_names <- colnames(x)  equation <- paste("last\_evaluation =", round(intercept, 3), "+", paste(round(coefficients, 3), variable\_names, collapse=" + "), "+ ε")  print(equation)  #Оценка модели при помощи коэффицента детерминации  y\_predicted <- predict (final\_ridge\_model, s = lambda\_optimal, newx = x)  sst <- sum ((y - mean (y))^2)  sse <- sum ((y\_predicted - y)^2)  rsq <- 1 - sse/sst  rsq  #График рассеяния для сравнения реальных и прогнозируемых значений  scatter\_plot\_predicted <- ggplot(data, aes(x = y, y = y\_predicted)) +  geom\_point() +  geom\_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") + # добавляем линию идеального соответствия  labs(x = "Реальный уровень удовлетворенности", y = "Прогнозируемый уровень удовлетворенности") +  ggtitle("Соответствие реальных и прогнозируемых уровней")  print(scatter\_plot\_predicted) |