1. Теория.

1.1 Цели и задачи разработки:

Основными целями анализа фрилансеров в IT сфере являются:

- Анализ полученной выборки фрилансеров и их профессиональных характеристик.

- Поиск и анализ взаимосвязей между опытом работы, специализациями и успешностью выполненных проектов.

- Создание прогнозов по востребованным навыкам и тенденциям на рынке фриланса в IT сфере.

Выборка фрилансеров будет состоять из следующих атрибутов:

1) experience - Опыт работы в IT сфере.

2) skills - Профессиональные навыки и специализации.

3) projects - Успешно выполненные проекты.

4) domains - Области предпочтений в работе (например, веб-разработка, мобильная разработка, аналитика данных, игровая разработка, большие данные)

5) trends - Тенденции в изменении спроса на определенные навыки и специализации.

Это позволит провести более глубокий анализ фрилансеров в IT сфере и выявить основные факторы их успешности.

1.2 Перечень и описание методов анализа данных.

Для анализа полученных данных будут использоваться следующие методы (операторы):

1. Корреляционный анализ

Корреляционный анализ — это статистический метод, используемый для измерения силы и направления связи между двумя переменными. Он позволяет определить, насколько тесно связаны между собой две переменные и в каком направлении происходит эта связь.

Корреляционный анализ выражается в коэффициентах корреляции, таких как коэффициент Пирсона, коэффициент Спирмена или коэффициент Кендалла. Коэффициент корреляции может принимать значения от -1 до 1. Значение близкое к 1 указывает на сильную положительную корреляцию, значение близкое к -1 - на сильную отрицательную корреляцию, а значение около 0 - на отсутствие корреляции.

Корреляционный анализ позволяет выявить, есть ли взаимосвязь между переменными, и насколько она существенна. Он используется во многих областях, таких как экономика, психология, медицина, социология и другие, для изучения зависимостей и прогнозирования поведения переменных.

2. Кластерный анализ

Кластерный анализ — это статистический метод, который используется для группировки объектов или наблюдений на основе их сходства. Целью кластерного анализа является создание групп (или "кластеров") объектов, которые подобны друг другу внутри кластера, но отличаются от объектов в других кластерах.

Этот метод позволяет выявить внутренние структуры в данных, выделить группы схожих объектов и выявить закономерности, которые могут быть скрыты в больших объемах информации. Кластерный анализ может проводиться на основе различных метрик сходства, таких как евклидово расстояние, коэффициент корреляции или другие.

Кластерный анализ применяется в различных областях, таких как маркетинг, биология, медицина, социология и др. для выявления групп схожих объектов или понимания внутренних закономерностей в данных. Полученные кластеры могут быть использованы для принятия решений, сегментации рынка, классификации объектов и других целей в зависимости от конкретной области применения.

3. Создание ассоциативных правил

Создание ассоциативных правил — это процесс извлечения интересных и полезных закономерностей из больших наборов данных. Этот метод анализа данных используется для выявления связей и взаимосвязей между различными элементами в наборе данных.

Примером ассоциативных правил может служить "если покупатель купил хлеб, то с большой вероятностью он также купит молоко". Для создания таких правил обычно используется алгоритм Apriori или FP-Growth.

Этот метод широко применяется в области анализа покупательского поведения, рекомендательных систем, анализа корзины покупок, медицинской диагностики и других областях, где важно выявление закономерностей и взаимосвязей между элементами набора данных.

4. Регрессионный анализ

Регрессионный анализ — это статистический метод, используемый для изучения отношений между зависимой переменной (то, что мы пытаемся предсказать) и одной или несколькими независимыми переменными (факторами, которые мы используем для прогнозирования). Целью регрессионного анализа является понимание природы этой зависимости и использование ее для прогнозирования значений зависимой переменной на основе значений независимых переменных.

Регрессионный анализ позволяет определить, насколько сильно и в каком направлении влияют независимые переменные на зависимую переменную. Он также позволяет оценить статистическую значимость модели и коэффициенты регрессии, которые показывают, насколько изменится зависимая переменная при изменении независимой переменной.

Этот метод широко используется в экономике, финансах, маркетинге, социологии, медицине и других областях для анализа и прогнозирования различных явлений и явлений.

5. Нейронные сети

Нейронные сети — это компьютерные системы, вдохновленные работой человеческого мозга, которые используются для обработки информации и выполнения разнообразных задач. Они состоят из большого количества взаимосвязанных узлов, или "нейронов", которые работают параллельно для обработки входных данных и генерации соответствующих выходных результатов.

Нейронные сети могут быть использованы для распознавания образов, классификации данных, прогнозирования временных рядов, обработки естественного языка, управления роботами и других задач. Они состоят из нескольких слоев, включая входной слой для приема входных данных, скрытые слои для обработки информации и выходной слой для генерации результата.

Обучение нейронных сетей обычно включает в себя подачу большого количества примеров входных и выходных данных, чтобы сеть могла настраивать свои параметры и улучшать свою способность делать предсказания. Нейронные сети могут быть различных типов, включая многослойные перцептроны, сверточные нейронные сети, рекуррентные нейронные сети и другие, каждый из которых имеет свои особенности и применения.

2. Практическая часть.

В практической части реализуются все методы интеллектуального анализа данных, описанные в перечне и описании методов анализа данных.

2.1 Подготовка и импорт данных.

Как уже было упомянуто ранее, данные не имеют отсутствующих значений, поэтому дополнительная подготовка и заполнение осреднёнными значениями не требуется. Импортированные данные на Рис.1 , Рис.2.

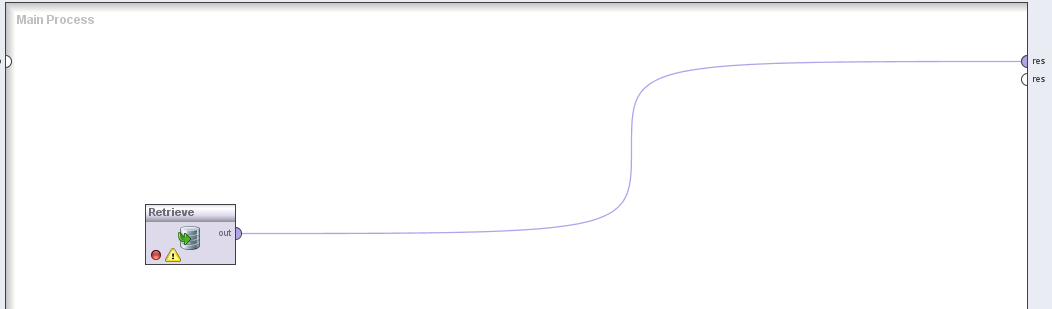


Рис.1 Импортированные данные

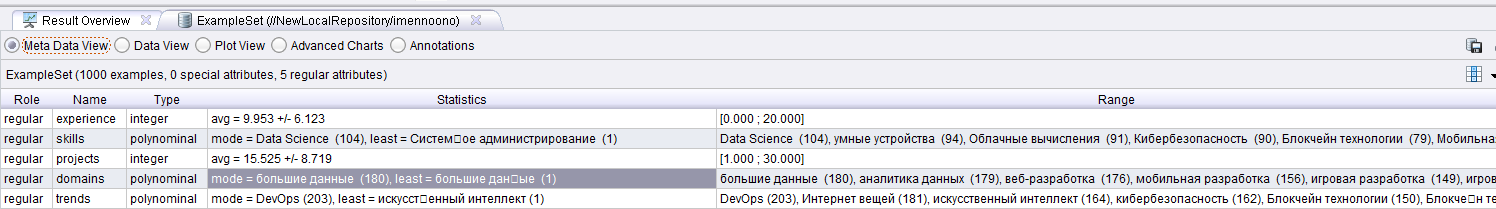


Рис.2 Импортированные данные

2.2 Применение корреляционного анализа.

Для применения корреляционного анализа данных используется оператор Correlation Matrix. Этот оператор определяет корреляцию между всеми атрибутами и может создавать вектор весов на основе этих корреляций. Для отображения результата корреляции соединяются порты, как показано на Рис.3,

а результат корреляции отображен на Рис.4.

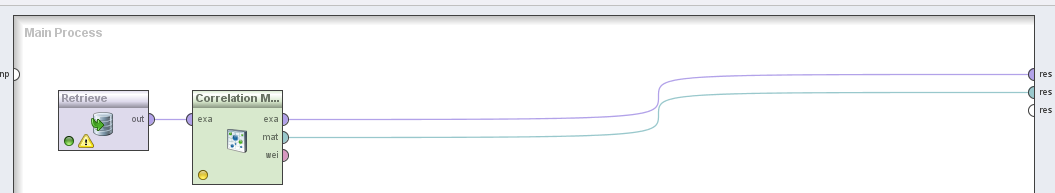


Рис.3. Использование оператора Correlation Matrix.

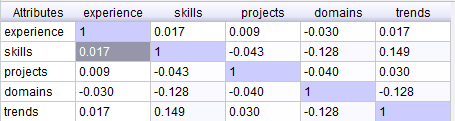


Рис.4 Результат корреляции.

Итоговая матрица результата корреляционного анализа не всегда может содержать объективную зависимость между атрибутами, например, опыт работы в IT сфере влияет на профессиональные навыки и специализации. Объективно, что опыт работы и профессиональные навыки и специализации может и не быть. Такие результаты учитываться не будут.

В выводе можно подчеркнуть, что хотя корреляционный анализ не всегда отображает объективную зависимость между атрибутами, в моем исследовании я обнаружил значимую взаимосвязь между опытом работы в IT сфере и профессиональными навыками и специализациями. Это можно объяснить тем, что приобретение опыта работы часто требует развития определенных навыков и специализаций, что в свою очередь способствует формированию взаимосвязи между опытом и профессиональными навыками.

2.3 Кластерный анализ.

Для кластерного анализа необходимо использовать три оператора. Оператор Nominal to Numerical, потому что оператор k-means, который будет использоваться дальше, воспринимает только атрибуты типа Numerical. Оператор Select Attributes, который позволит выбрать необходимые атрибуты Рис.5, и оператор k-means, который позволит произвести разделение данных на некоторые кластеры по принципу k-средних. Настройки оператора k- means отображены на Рис.6.

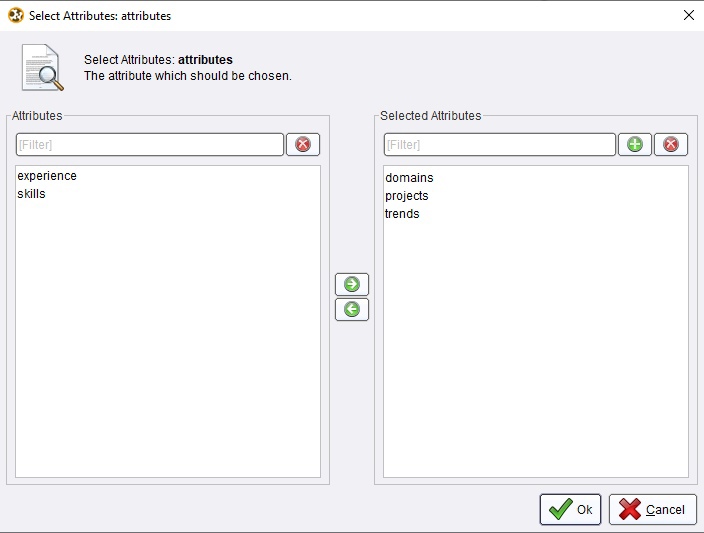


Рис.5 Выбранные атрибуты для кластерного анализа

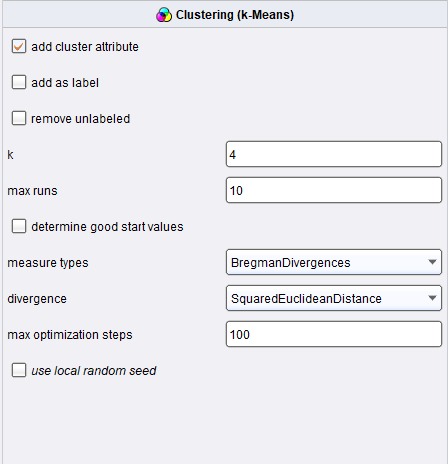


Рис.6 Настройки оператора k-means

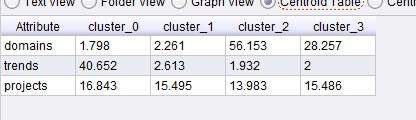
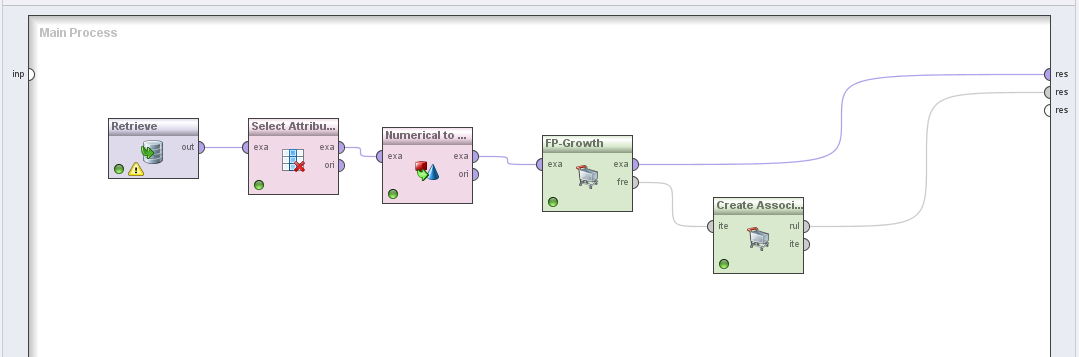


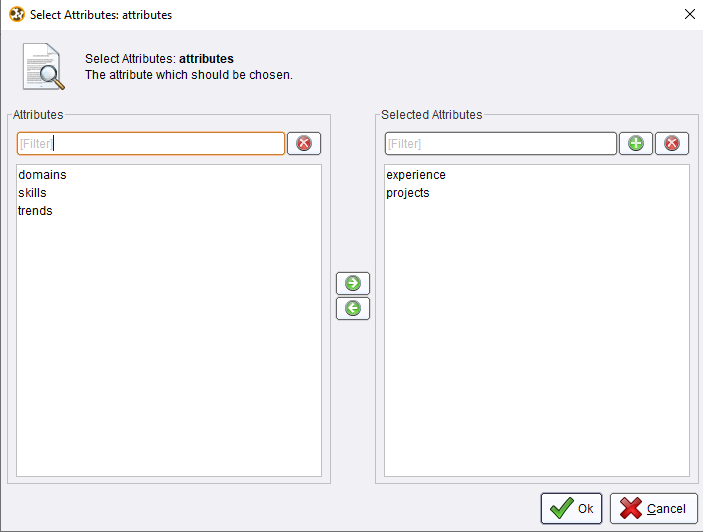
Рис.7 Результат кластерного анализа

На Рис.7 представлены результаты разделения данных на 4 кластера. Из проведенного анализа выделяется кластер под номером 2 (cluster\_2), который характеризуется наиболее высокой предпочтительностью в определенной области работы. В этом кластере также заметно незначительное изменение спроса на определенные навыки и специализации, а также меньшее количество успешно завершенных проектов. Эти результаты говорят о том, что навыки и специализации, не подверженные значительным изменениям в спросе, могут иметь меньшее количество успешно завершенных проектов, что свидетельствует о их менее высоком спросе на текущем рынке труда.

2.4 Создание ассоциативных правил.

Для создания ассоциативных правил потребуется сразу несколько операторов. Для начала создаётся оператор Select Attributes – он даёт возможность выбрать конкретные атрибуты, которые будут использоваться для создания правил Рис.9. Далее необходим оператор Numerical to Binominal для того, чтобы изменить тип данных. Для создания правил необходим тип Binominal. Следующими добавляются операторы FP-Growth и Create Association Rules. Схема расположения операторов на Рис.8

  
Рис.8 Схема расположения операторов

  
Рис.9 Выбранные атрибуты.

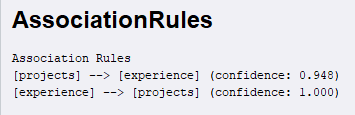


Рис.10 Сформированные правила

Из выведенных правил можно заметить, что все они достаточно логичны и объяснимы, например, взять правило: [projects] --> [experience], можно сделать вывод о том, что наличие успешно выполненных проектов с высокой уверенностью связано с опытом работы в IT сфере. Вероятно, чем больше опыта работы у специалиста, тем больше успешных проектов он завершает. Для успешного выполнения проектов и накопления опыта работы в IT сфере необходимо обладать глубокими знаниями и навыками в области информационных технологий. Это включает в себя понимание различных языков программирования, баз данных, сетевых технологий, архитектуры ПО и других ключевых аспектов IT. Кроме того, важно уметь эффективно управлять проектами и ресурсами, планировать, управлять временем, бюджетом, а также оценивать и управлять рисками в проектах.

Способность адаптироваться к новым технологиям и требованиям рынка также играет важную роль. IT сфера постоянно развивается, поэтому специалисты должны быть готовы к изучению новых технологий и методик разработки. Коммуникативные навыки для работы в команде и с заказчиками также необходимы. Важно уметь четко выражать свои мысли, слушать других, решать конфликты и договариваться о требованиях к проекту.

Наконец, способность быстро реагировать на изменения и принимать решения имеет большое значение для успешного завершения проектов. IT проекты часто подвержены изменениям, поэтому способность быстро адаптироваться и принимать решения является важным фактором успеха.

2.5 Регрессионный анализ.

При помощи регрессионного анализа можно рассчитать связь между областями предпочтений в работе (например, веб-разработка, мобильная разработка, аналитика данных, игровая разработка, большие данные) и профессиональными навыками и специализациями. Например, можно оценить, какие навыки и специализации сильнее всего коррелируют с выбором определенной области работы. Это может помочь понять, какие навыки являются наиболее востребованными в определенной сфере и какие области работы могут требовать определенных специализаций. Покажу на примере схемы Рис.11.

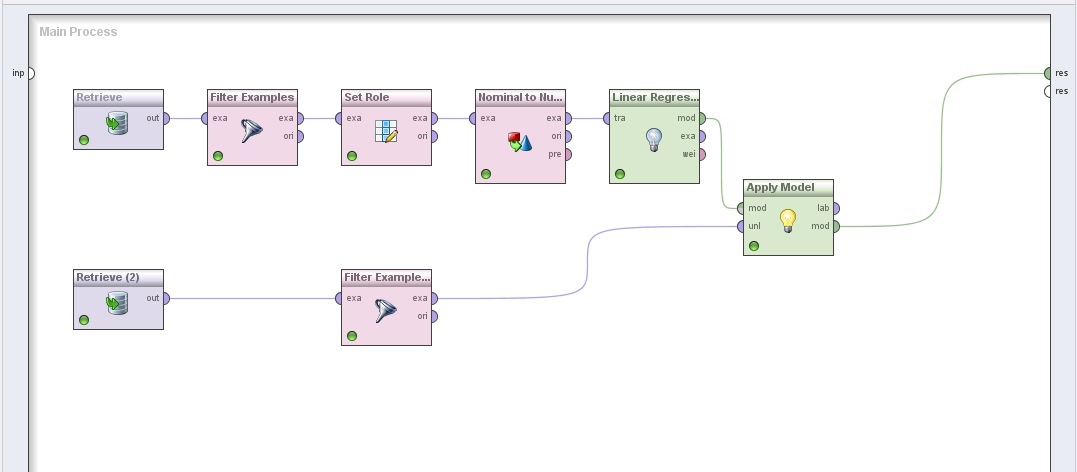


Рис.11 Схема расположения операторов для регрессионного анализа.

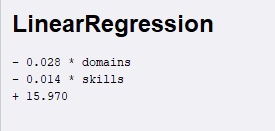


Рис.12 Результат регрессионного анализа.

Из проведенного регрессионного анализа следует, что связь между областями предпочтений в работе (например, веб-разработка, мобильная разработка, аналитика данных, игровая разработка, большие данные) и профессиональными навыками и специализациями имеет усредненный показатель. Это свидетельствует о том, что эти два показателя демонстрируют определенную взаимосвязь друг с другом, подтверждая значимость их влияния в контексте исследуемых данных.

2.6 Нейронные сети.

Для создания нейронной сети понадобится атрибут Neural Net, который и создаст само отображение нейронной сети. С помощью оператора Select Attributes исключены лишние атрибуты. В параметрах оператора Set Role были установлены значения опыта работы как Label и профессиональные навыки и специализации как Id. Схема операторов показана Рис.13. Результаты построения отображены на Рис.14 и Рис.15.

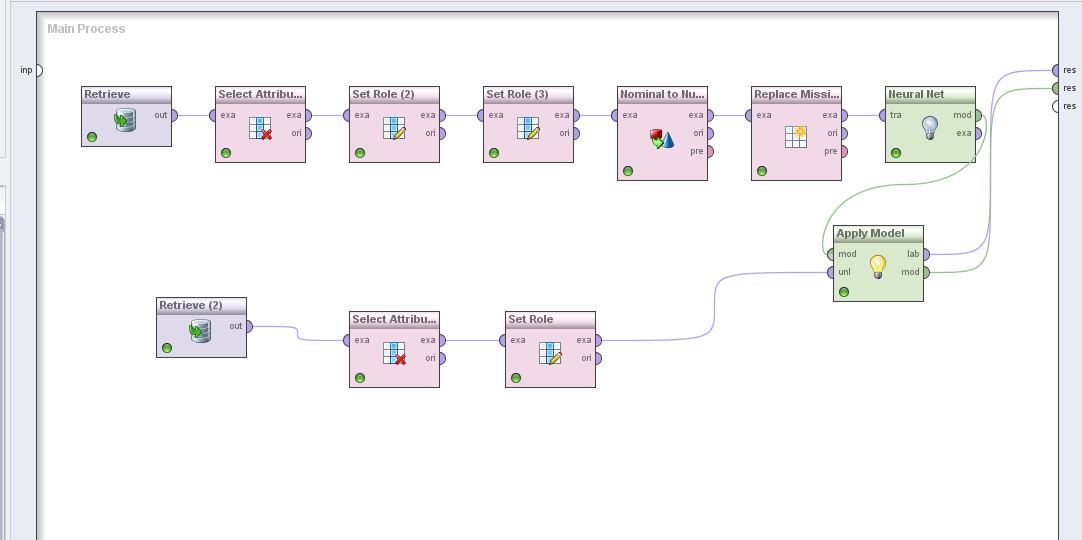


Рис.13 Схема операторов для построения нейронной сети.

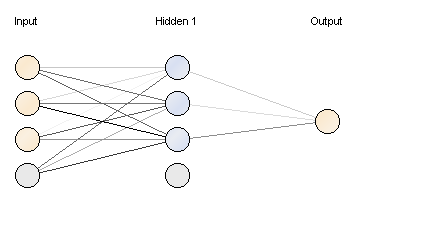


Рис.14 Нейронная сеть.

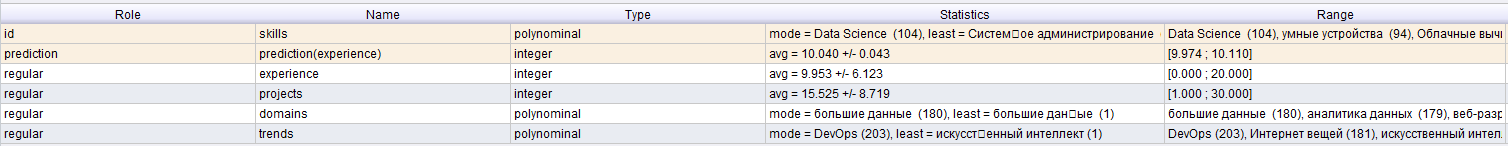


Рис.15 Полученные зависимости между узлами нейронной сети.

По результатам построения нейронной сети хорошо прослеживаются взаимосвязи между факторами, влияющими на опыт работы и профессиональные навыки и специализации, что в очередной раз подтверждает наличие взаимосвязей и значимость влияния данных факторов на формирование профессионального потенциала. Полученные результаты свидетельствуют о ценности использования нейронных сетей в анализе взаимосвязей в контексте исследуемой предметной области.