

ПРОГНОЗИРОВАНИЕ ЗАРАБОТНОЙ ПЛАТЫ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ МЕТОДОВ

Н.В. Шкляева

*Новосибирский государственный технический университет
Новосибирск*

В современных реалиях рынок труда является одним из показателей, по состоянию которого можно судить о благополучии, стабильности, эффективности политики и развития регионов и страны в целом. А быстрый рост использования интернета для поиска и размещения предложений по трудоустройству предоставляет прекрасную возможность для мониторинга рынка труда в режиме реального времени. Анализ вакансий в Интернете, действительно, представляет конкурентное преимущество для участников рынка труда в отношении классического анализа на основе опросов. Поскольку он является богатым источником информации для понимания динамики и тенденций рынка труда.

Все это способствовало появлению множеству различных исследований, в которых авторы изучали различные подходы к анализу рынка труда и выявления ключевых факторов, оказывающих наибольшее влияние [1]. Одним из этих факторов является зарплата, изучение которой будет также интересно и для большей части населения.

Знания человека о собственной ценности, не только с точки зрения компании, но и с точки зрения рынка труда, будет полезна при поиске работы, при продвижении по карьерной лестнице внутри компании и при переходе в другую компанию.

Помимо прочего на рынке труда существует проблема непонимания реальной ценности конкретных компетенций. Для соискателя это может привести к неправильно выстроенным зарплатным и должностным ожиданиям. А обладая информацией о ценности компетенций, можно не только правильно написать резюме, но также понять какие навыки стоит развить в себе для повышения собственной ценности.

Компаниям также необходимо знать и учитывать тенденции рынка, так как отсутствие такого анализа может повысить вероятность оттока ценных кадров, снижения мотивации сотрудников и т.д.

Исходя из выше сказанного, задача данного исследования состоит в том, чтобы определить, как компетенции влияют на уровень зарплаты. Для решения этой задачи воспользуемся самым популярным методом для прогнозирования на основе онлайн-данных - нейросетями.

Нейронные сети - это очень мощный и гибкий механизм прогнозирования. В связи с этим нейросетевое прогнозирование является одной из наиболее интересных и имеющих практическое применение областей использования нейросетевых подходов.

Важнейшим преимуществом нейронных сетей с точки зрения прогнозирования заключается в отсутствии строгой формализации. Все это дает широкий спектр возможностей исследования, усовершенствования и адаптации существующих моделей нейронных сетей и повышения точности прогноза.

Построение нейросетевой системы включает в себя обработку входных данных, разработку архитектуры и обучение сети. На данный момент не существует единого алгоритма реализации каждого из перечисленных этапов, так как конфигурация системы зависит от множества факторов связанных с конкретной задачей. Поэтому при разработке нейронной сети важно учитывать природу исходных данных, объем входных данных и т.д. [2].

Искусственная нейросеть представляет собой совокупность нейронов, взаимодействующих друг с другом, в результате чего происходит обработка информации и обмен данными между собой. Обмен данными происходит через синаптические связи, соединяющие выход одного нейрона с входом другого. Каждая связь характеризуется весовым коэффициентом, благодаря которому, входная информация изменяется, когда передается от одного нейрона к другому. По мере того как происходит обработка информации, нейрон суммирует входные сигналы, вычисляет от полученной суммы функцию и после чего передает результирующее значение на выход [3].

Существуют различные виды искусственных нейронных сетей, которые отличаются друг от друга строением и процессом обучения. Однако наиболее распространенной и успешно применяемой в прогнозировании является многослойный перцептрон. Эта сеть, в которой помимо входного и выходного слоя присутствуют и скрытые слои нейронов, число которых зависит от решаемой задачи.

Многослойный перцептрон относится к классу сетей, обучение которых осуществляется методом обратного распространения ошибки [4]. Для обучения системы используется обучающая выборка, состоящая из обучающих примеров, формирование которых основывается на исходных данных и специфике решаемой задачи.

При решении задачи прогнозирования нейросетевая система строится следующим образом: входной слой содержит несколько нейронов, на которые подаются значения исследуемого временного ряда, а последний слой состоит из единственного нейрона, на выходе которого получается прогноз.

Для построения нейросети будем использовать программу аналитической платформы Deductor. Нейронная сеть представляет собой многослойный перцептрон с одним скрытым слоем из 10 нейронов, так как экспериментально было выявлено, что такое количество обладает высокой скоростью обучения и при этом имеет низкий уровень ошибки для исследуемого набора данных.

Количество входных нейронов в сети – 9, что соответствует количеству входных параметров – компетенций претендентов на должность в сфере программирования и разработки (наиболее часто встречаемые в объявлениях в интернет-бирже труда – HeadHunter требования к кандидатам):

- Опыт использования реляционных СУБД или NoSQL-хранилищ;
- Опыт работы в команде с использованием гибких методологий;
- Уверенные знания, понимание принципов ООП и паттернов;
- Знание английского языка;
- Высшее образование;
- Знание нескольких языков программирования;
- Опыт в коммерческих и промышленных разработках;
- Опыт работы в Linux из командной строки;
- Опыт работы.

Входной параметр такой как «Опыт работы» принимает оценку данного навыка от 0 до 4, где:

- без опыта работы,
- требуемый опыт от 1 года до 3 лет,
- требуемый опыт от 3 до 6 лет,
- требуемый опыт более 6 лет.

Все остальные входные параметры принимают значение 1, если такой навык требуется от кандидата, в противном случае - 0.

Поскольку в регионах наблюдается значительные различия в уровне оплаты труда то, было принято решение анализировать только один регион, в связи с чем исследование проводилось только по вакансиям Новосибирской области.

Описанные выше данные были получены путем обращения к публичному API сайта HeadHunter, в результате чего для исследования было собрано 1334 вакансий. Обязательным условием при сборе данных было наличие заработной платы в вакансии. Для выявления ключевых навыков из объявления был использован текстовый анализ.

Так как в объявлениях чаще всего указывают диапазон заработной платы, то на выходе у сети будем использовать 2 нейрона, которые соответствуют началу диапазона прогнозируемых зарплат и концу данного диапазона. В ходе многочисленных испытаний, было выявлено, что наиболее удачной функцией активации (нелинейная функция, которая вычисляет выходной сигнал формального нейрона) для исследуемых данных является сигмоида. А для обучения сети лучше всего использовать метод Resilient Propagation (Rprop) [2].

В результате мы получили многослойную нейронную сеть, которая способна прогнозировать диапазон заработной платы в зависимости от требуемых навыков кандидата на должность с уровнем ошибки в 0,0049. На тестовой выборке средняя ошибка составляет 0,008, что также довольно неплохой результат.

Оценить качество построенной модели можно также и с помощью диаграммы рассеивания (рис. 1-2), на которой видно, что прогнозируемые значения имеют небольшое от реальных. Большая часть точек сосредоточена на небольшом расстоянии от линии идеальных оценок и находится в пределах заданного «коридора» погрешности.

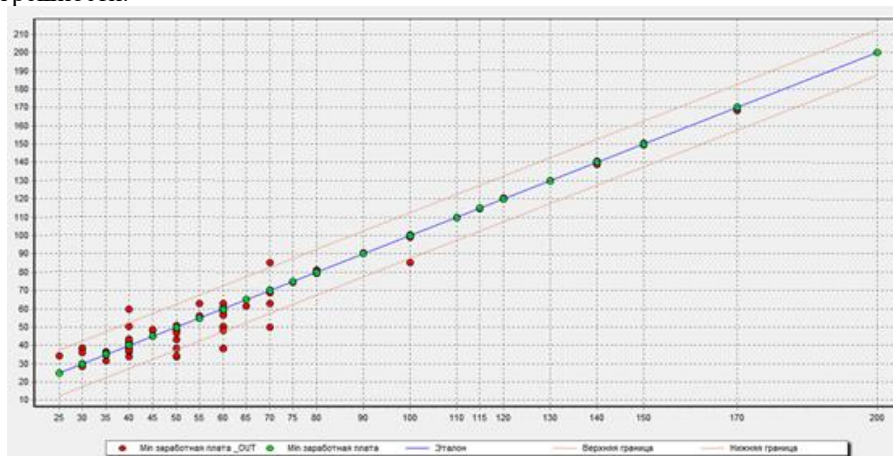


Рис. 1. Диаграмма рассеивания прогнозируемых значений от реальных для минимальной з/п (тыс. руб.)

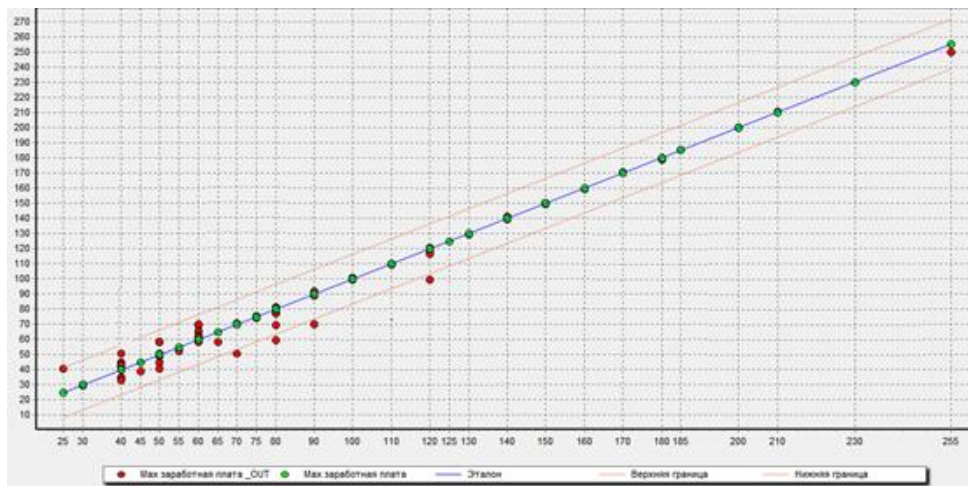


Рис. 2. Диаграмма рассеяния прогнозируемых значений от реальных для максимальной з/п (тыс. руб.)

С помощью поочередной активации входных нейронов и наблюдением за результатом сети можно определить степень влияния входных параметров модели на минимум и максимум диапазона прогнозируемой заработной платы. Наиболее значимыми компетенциями претендентов на должность в сфере программирования и разработки оказались «Знание нескольких языков программирования», «Опыт работы» и «Опыт в коммерческих и промышленных разработках». Менее значимыми навыками являются «Опыт работы в команде с использованием гибких методологий», «Опыт использования реляционных СУБД или NoSQL-хранилищ». На рисунке 3 представлена гистограмма значимости входных параметров на выходные значения.

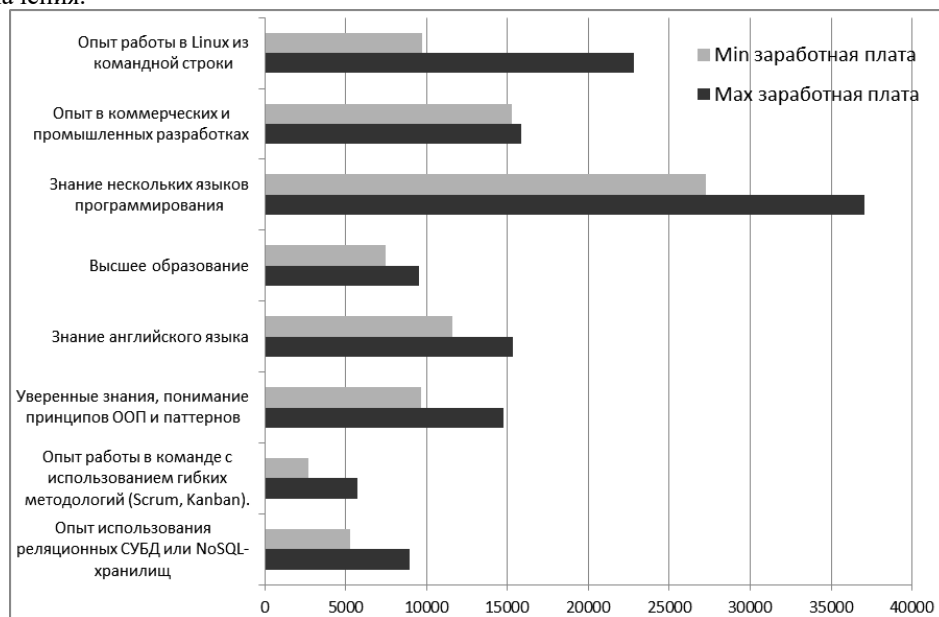


Рис. 3. Значимость входных параметров

Высокая значимость такого навыка, как «Знание нескольких языков программирования» объясняется тем, что все чаще работодатели хотят видеть «универсальных» сотрудников, способных решать различные задачи. «Опыт работы» и «Опыт в коммерческих и промышленных разработках» также обладают высокой ценностью, так как являются доказательством применения своих умений и возможностей на практике соискателем на должность.

Несмотря на то, что полученная модель обладает достаточно высокой точностью, у нее имеется недостаток. Так как данные были взяты из вакансий, в которых обычно требуется определенный специалист с определенным набором навыков в рамках какой-нибудь конкретной экосистемы для решения какой-либо конкретной задачи. И в этих случаях заработная плата будет формировать не только исходя из навыков, но и исходя из тех задач и типа проекта, на которые идет специалист, что невозможно учесть при прогнозировании.

ЛИТЕРАТУРА

1. Кобец Е. А. Факторы, влияющие на рынок труда // Инновационная наука. 2016. №. 8-1. С. 69-71.
2. Boselli R. et al. Labour Market Intelligence for Supporting Decision Making // SEBD. 2017. С. 74.
3. Хайкин С. Нейронные сети: полный курс / Саймон Хайкин. // М.: Издательский дом «Вильямс», 2006. 104 с.
4. Saputra W. et al. Analysis resilient algorithm on artificial neural network backpropagation // Journal of Physics: Conference Series. – IOP Publishing, 2017. Т. 930. №. 1. С. 12-35.

АННОТАЦИЯ НА РУССКОМ ЯЗЫКЕ

В статье описано применение нейронной сети для прогнозирования диапазона заработной платы специалистов в сфере программирования и разработки для Новосибирской области. В качестве инструмент анализа данных предложено использовать аналитическую платформу Deductor. Представлена основная информация об архитектуре нейронной сети, а также определены наиболее значимые параметры, оказывающие наибольшее влияние на ожидаемую зарплату.

Ключевые слова на русском языке: нейронная сеть, прогнозирование, заработная плата, интернет.

SALARY FORECASTING USING NEURAL NETWORK METHODS

N.V. Shklyueva

Novosibirsk State Technical University

Novosibirsk

The article describes the use of a neural network to predict the salary range of specialists in the field of programming and development for the Novosibirsk region. It was proposed to use the Deductor analytical platform as a data analysis tool. Basic information about the architecture of the neural network is presented, and the most significant parameters that have the greatest impact on the expected salary are identified.

Keywords in English: neural network, forecasting, salary, Internet.