ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Научная статья УДК 004.855.5 EDN UHRDFB

https://doi.org/10.34216/2587-6147-2024-2-64-51-58

Дмитрий Михайлович Григорьев¹ Мария Алексеевна Григорьева²

^{1,2}Костромской государственный университет, г. Кострома, Россия ¹ generalgrigorevous@gmail.com, https://orcid.org/0009-0001-2636-561X

ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПРИ АНАЛИЗЕ СООТВЕТСТВИЯ СОИСКАТЕЛЯ РАССМАТРИВАЕМОЙ ДОЛЖНОСТИ

Аннотация. Возможно ли спрогнозировать долгосрочность трудовых отношений и корректно оценить целесообразность вложения материальных средств и нематериальных благ в кандидата на должность? Возможно ли отказаться от психологического аспекта при рассмотрении соискателя и сделать упор исключительно на математических и логических показателях, указанных в его резюме? Данная статья рассматривает возможность использования математических методов при первичном анализе соответствия кандидата требованиям должности. Статья затрагивает исключительно математический аспект, т. е. предлагает методику расчета рейтинга соискателя на основании сравнения ключевых показателей из его резюме с требованиями должности с использованием метода анализа иерархий и вычисления косинусного расстояния между текстами с целью дальнейшего определения корреляции между компетенциями соискателя и навыками, необходимыми для успешного трудоустройства на должность. В работе приведены частные статистические данные и примеры использования указанной методики на тестовых резюме.

Ключевые слова: метод анализа иерархий, косинусное расстояние, косинусное сходство, математические модели, прикладное использование математических моделей, управление человеческими ресурсами, автоматизация управления человеческими ресурсами, соискатель, оценка соискателя, вакансия, требования вакансии

Для цитирования: Григорьев Д. М., Григорьева М. А. Использование математических моделей при анализе соответствия соискателя рассматриваемой должности // Технологии и качество. 2024. № 2(64). С. 51-58. https://doi.org/10.34216/2587-6147-2024-2-64-51-58.

Original article **Dmitrij M. Grigor'ev**¹ **Mariya A. Grigor'eva**²

^{1,2}Kostroma State University, Kostroma, Russia

THE USAGE OF MATHEMATICAL METHODS IN ANALYSIS OF SUITABILITY OF AN APPLICANT FOR THE CONSIDERED POSITION

Abstract. Is it possible to predict the long-term nature of labor relations and correctly assess the expediency of investing material resources and non-material benefits in a candidate for a position? Is it possible to abandon the psychological aspect when considering an applicant and focus only on the mathematical and logical indicators of his resume? This article considers the possibility of usage of mathematical methods in the primary analysis of a candidate's compliance with the requirements of the position. Current article deals exclusively with the mathematical aspect and proposes a methodology for calculating the applicant's rating based on comparing key indicators of his resume with the requirements of the position using the analytic

© Григорьев Д. М., Григорьева М. А., 2024

_

² themarymakii@gmail.com, https://orcid.org/0009-0009-2250-3948

hierarchy process and calculating the cosine distance between texts in order to further determine the correlation between the applicant's competencies and skills, which are necessary for a successful employment in the position. This article presents specific statistical data and examples of the usage of this methodology on test resumes.

Keywords: analytic hierarchy process, cosine distance, cosine similarity, mathematical methods, an application of mathematical methods, HR, HR-management, automation of HR-management, applicant, analysis of an applicant, vacancy, vacancy requirements

For citation: Grigor'ev D. M., Grigor'eva M. A. The usage of mathematical methods in analysis of suitability of an applicant for the considered position. Technologies & Quality. 2024. No 2(64). P. 51–58. (In Russ.). https://doi.org/10.34216/2587-6147-2024-2-64-51-58.

Развитие организации любой направленности тесно связано с персоналом, т. е. с людьми, которые работают в данной организации, выполняют разного рода задачи, реализуют разной сложности проекты и в конечном итоге приносят организации прибыль. Актуально это как для государственных организаций, так и для частных. Однако для вторых борьба за человеческие ресурсы является более интенсивной, так как частные предприятия обычно не зависят от бюджетов, выделяемых выписстоящими инстанциями, и могут позволить себе самостоятельно устанавливать как условия приема соискателей, так и размер их вознаграждений, будь то непосредственно заработная плата или прочие социальные блага.

Однако, как понять, что соискатель стоит средств, которые в него предполагается вложить? Как понять, что будущий сотрудник принесет компании прибыль, а не убытки? Это сложные вопросы, на которые нет однозначного ответа, так как все случаи являются слишком частными для создания единых формул успешного найма. Тем не менее на основании частных случаев можно сформировать статистику и оценить зависимости между первичными ожиданиями соискателей, их предыдущим опытом работы и долгосрочностью будущих трудовых отношений [1, 2]. На основании указанной статистики можно рассматривать использование методов автоматизации отбора соискателей при поиске сотрудников.

В данном контексте было проведено исследование, направленное на определение зависимости между ожиданиями кандидата на должность, требованиями работодателя и целесообразностью трудоустройства кандидата. В основе исследования лежали следующие инструменты:

- 1) обучение и использование модели случайного леса;
- 2) использование метода анализа иерархии при оценке альтернативы на основании заданных параметров;
- 3) определение косинусного сходства между реальными навыками соискателя и навыками, требуемыми для трудоустройства на должность.

Для практической реализации первого инструмента и его дальнейшего применения на практике были использованы наборы данных с Kaggle: датасет I, включающий в себя свыше 70 000 записей о соискателях, их уровне образования, опыте работы, зарплатных ожиданиях и степени соответствия навыков, и датасет II, содержащий более 4000 записей об уровне образования кандидатов, релевантном опыте и уровне заработной платы («низкий», «средний», «высокий»). С обоими датасетами была проведена предварительная работа: уровни образования были переведены из строкового типа в целочисленный на основании выделенных образовательных ступеней (табл. 1), а опыт работы был переведен в месяцы с целью получения более точных результатов при дальнейшем обучении.

Таблица 1 Образовательные ступени

Уровень образования / Ученая степень	Ступень
Среднее образование	1
Неоконченное высшее	2
Высшее образование	3
Магистр	4
Специалист	4
Кандидат наук	5
Доктор наук	6

Ключевой выходной параметр датасета І Employed является логической переменной и содержит информацию о том, был ли кандидат принят на работу. Входными параметрами являются EdLevel (уровень образования), Mental Health (состояние психологического здоровья), TotalExperience (трудовой стаж в месяцах), AverageExperience (средняя продолжительность трудовых отношений с одним работодателем), Country (страна соискателя), SalaryExpectations (зарплатные ожидания) и SkillsSimilarity (степень соответствия навыков соискателя требуемым компетенциям). По этой причине датасет I был использован для обучения модели, роль которой сводится к прогнозированию целесообразности рассмотрения соискателя в качестве потенциального сотрудника.

Датасет I был нормализован, т. е. все некорректные значения были убраны, строчные значения были переведены в числовые (изменения коснулись входящего параметра EdLevel), а параметры MentalHealth и Country, которые не поддаются измерению, были исключены из рассмотрения. Также датасет I был сбалансирован, после чего использован для обучения модели, роль которой сводится к прогнозированию целесообразности рассмотрения соискателя в качестве потенциального сотрудника.

На рис. 1 продемонстрирована выборка из исходного состояния датасета І. На рис. 2 показана выборка из обработанного датасета І.

	EdLevel	MentalHealth	TotalExperience	AverageExperience	Country	SalaryExpectations	SkillsSimilarity	Employed
0	Master	No	86	22	Sweden	51000	0,4	0
1	Undergraduate	No	144	36	Spain	46000	1,2	1
2	Master	No	188	63	Germany	77000	0,7	0
3	Undergraduate	No	109	36	Canada	46000	1,3	0
4	PhD	No	485	162	Singapore	160000	0,2	0
5	Master	No	114	114	France	38000	0,5	0
6	Master	No	320	320	Germany	77000	1,7	1
7	Master	No	178	59	Switzerla	81000	0,4	0
8	Undergraduate	No	473	237	United Ki	68000	0,3	0
9	Master	No	243	61	Russian F	37000	0,6	0
10	Undergraduate	Yes	55	14	Israel	122000	1,8	1
11	Undergraduate	Yes	80	16	Turkey	11000	1,3	1
12	Master	No	232	232	Germany	60000	0,5	0
13	Undergraduate	No	96	24	United Sta	103000	0,7	0
14	Master	No	83	28	France	25000	0,5	0
15	Undergraduate	Yes	269	54	Brazil	60000	1,6	1
16	Other	Yes	87	29	Bulgaria	20000	0,5	0
17	Undergraduate	No	152	30	Greece	25000	2,5	1
18	Master	Yes	414	69	United Ki	64000	1,4	1
19	Master	No	260	130	Italy	54000	1,7	1
20	Other	No	62	31	Russian F	22000	1,2	0
21	Other	Yes	306	153	Canada	71000	0,6	0
22	Master	No	251	251	Spain	58000	0,7	1
23	Undergraduate	No	291	146	United Ki	129000	1,2	0
24	Master	No	126	126	Netherlar	25000	0,7	0
25	Undergraduate	No	308	308	United Sta	105000	0,6	0
26	Undergraduate	No	154	77	Poland	45000	3,9	1
27	PhD	No	129	22	Canada	151000	0,8	C
28	Undergraduate	No	129	129	Sweden	40000	0,8	C
29	Undergraduate	No	149	30	Switzerla	43000	2,4	1
30	Master	No	302	151	Hungary	42000	2,1	1
31	Other	No	378	126	Canada	95000	1	1
32	PhD	No	225	225	United Sta	100000	0,8	C
33	Undergraduate	No	321	321	United Kir	61000	1	1

Рис. 1. Выборка из датасета І в исходном состоянии

	Education	TotalExperience	Average Experience Salary Expectation	ns SkillsSimilarity	Hired
0	4	86	22 51	000 0,4	0
1	3	144	36 46	000 1,2	2 1
2	4	188	63 77	000 0,7	7 0
3	3	109	36 46	000 1,3	0
4	6	485	162 160	000 0,2	0
5	4	114	114 38	000 0,5	0
6	4	320	320 77	000 1,7	7 1
7	4	178	59 81	000 0,4	0
8	3	473	237 68	000 0,3	0
9	4	243	61 37	000 0,6	0
10	3	55	14 122	000 1,8	3 1
11	3	80	16 11	000 1,3	1
12	4	232	232 60	000 0,5	0
13	3	96	24 103	000 0,7	7 0
14	4	83	28 25	000 0,5	0
15	3	269	54 60	000 1,6	5 1
16	1	87	29 20	000 0,5	0
17	3	152	30 25	000 2,5	5 1
18	4	414	69 64	000 1,4	1
19	4	260	130 54	000 1,7	7 1
20	1	62	31 22	000 1,2	0
21	1	306	153 71	000 0,6	0
22	4	251	251 58	000 0,7	7 1
23	3	291	146 129	000 1,2	0
24	4	126	126 25	000 0,7	7 0
25	3	308	308 105	000 0,6	0
26	3	154	77 45	000 3,9	1
27	6	129	22 151	3,0	0
28	3	129	129 40	3,0	0
29	3	149	30 43	000 2,4	1
30	4	302	151 42	000 2,1	1
31	1	378	126 95	000	1
32	5	225	225 100	000	0
33	3	321	321 61	000	1

Рис. 2. Выборка из датасета I после нормализации и балансировки

Ключевой выходной параметр датасета II LeaveOrNot также является логической переменной, но содержит информацию о том, покинет ли потенциальный сотрудник место работы в ближайшие два года. Входными параметрами являются Education (уровень образования), PaymentTier (уровень запроса кандидата по размеру заработной платы, где значение «1» говорит о запросе меньшем, чем готов предложить работодатель, значение «2» говорит о попадании в зарплатную вилку, предлагаемую работодателем, а значение «3» говорит о запросе, превышающем предложение работодателя), ExperienceMonth (трудовой стаж в месяцах), Average (средняя продолжительность трудовых отношений с одним работодателем), Extroversion (показатель экстраверсии от 1 до 10 по метрикам психологического тестирования BigFive), Neuroticism (показатель нейротизма от 1 до 10 по метрикам психологического тестирования BigFive), Kindness (показатель дружелюбия от 1 до 10 по метрикам психологического тестирования BigFive), Scrupulosity (показатель ответственности от 1 до 10 по метрикам психологического тестирования BigFive) и Intellect (показатель интеллекта от 1 до 10 по метрикам психологического тестирования BigFive).

Датасет II также был нормализован, т. е. все некорректные значения были убраны, а строчные значения были переведены в числовые (изменения коснулись входящего параметра Education). Также датасет II был сбалансирован, после чего использован для обучения другой модели, прогнозирующей приблизительную длительность трудовых отношений с потенциальным сотрудником.

На рис. 3 продемонстрирована выборка из исходного состояния датасета II. На рис. 4 показана выборка из обработанного датасета II.

Для применения математических методов (метод анализа иерархии и определение косинусного сходства) были рассмотрены критерии, выдвигаемые работодателем к соискателю при создании вакансий на агрегаторе hh.ru. В их число вошли:

- требования к уровню образования;
- требования к опыту работы [3, 4];
- предлагаемая заработная плата;
- требования к графику работы и форме занятости;
- требования по возможности переезда и командировок.

Дополнительным параметром стала надежность кандидата, вычисляемая как отношение общего стажа к количеству мест работы.

Первой математической моделью стал метод анализа иерархий, используемый при принятии решений при поиске альтернатив [5]. При проведении данного исследования метод был видоизменен с целью исключения сравнения кандидатов – теперь кандидаты должны были сравниваться только с требованиями вакансии. Каждому кандидату выставляется рейтинг, имеющий прямую зависимость от его соответствия параметрам вакансии. При этом параметры должны быть настраиваемыми, т. е. работодатель при создании вакансии должен иметь возможность установить ранги, повышающие или понижающие значимость требования при рассмотрении кандидата. Как было сказано ранее, решено было рассматривать шесть параметров, т. е. рангов тоже получилось шесть. При этом работодатель может указать одинаково высокий или одинаково низкий ранг у каждого критерия, что в итоге приведет к одному и тому же результату при вычислении весов параметров, влияющих на формирование рейтинга соискателя.

Следующим математическим методом стало вычисление косинусного расстояния между текстами с целью последующего определения их сходства [6], что позволило бы оценить соответствие навыков кандидата тем навыкам, которые перечислены в вакансии. С одной стороны, достаточно было определить вхождения одних навыков в массив других, но при перечислении компетенций могут менять формулировки, в связи с чем данное сравнение показало бы грубый результат, не соответствующий действительности. Таким образом, было решено принимать на вход два текста, приводить лингвистические единицы к леммам, определять общие массивы слов и затем определять совпадение. Чем выше уровень совпадения, тем выше соответствие компетенций кандидата навыкам, требуемым в вакансии.

Показатель косинусного сходства, равно как и стандартные требования, также был использован при определении соответствия соискателя. Таким образом, для вычисления рейтинга соискателя были разработаны формулы, приведенные в табл. 2.

На рис. 5 продемонстрированы требования должности. Данные требования являются тестовыми и предполагают исключительно демонстрацию возможностей разрабатываемой системы и применяемых инструментов.

Education P	aymentTier	ExperienceMonth	Average	Extroversion	Neuroticism	Kindness	Scrupulosity	Intellect	LeaveOrNot
0 Bachelors	3	4	2	2	. 1	. 9	8	10	
1 Bachelors	1	. 78	10	3	4	5	10	9	
2 Bachelors	3	29	6	8	9	3	10	10	
3 Masters	3	183	23	2	. 3	6	10	3	
4 Masters	3	56	14	. 6	6	5	10	5	
5 Bachelors	3	8	4		2	2 8	5	10	
6 Bachelors	3	9	9	3	3	3	1	6	
7 Bachelors	3	57	7	2	2	2 7	7	6	
8 Bachelors	3	34	17	•	1	10	10	8	
9 Masters	2	83	21		4	1	1	9	
LO Masters	3	180	26	8	7	7 2	1	3	
11 Bachelors	3	82	12	10	8	9	8	3	
12 Bachelors	3	61	. 6	2		3	3	5	
13 Bachelors	3	25	6	1	. 5	7	4	6	
L4 Bachelors	3	150	15		9	7	7	3	
L5 Bachelors	1	36	6	3	8	5	6	5	
L6 Bachelors	3	33	5	,	1	. 7	6	2	
17 Bachelors	3	98	12	2	. 5	8	1	7	
18 Bachelors	2		4		9	9	8	9	
19 Bachelors	2		4		7	7	9	6	
20 Bachelors	3	4	. 2	1	. 1	. 8	9	10	
21 Masters	2	148	21		. 9	8	8	7	
22 Bachelors	2	81	. 9	3	3	3	7	6	
23 Bachelors	3	62	. 8	9	9	10	4	5	
24 Bachelors	3	150	25	2	. 2	2 3	8	1	
25 Bachelors	3	4	4		4	1	3	1	
26 Bachelors	3	54	18	8	3	9	2	5	
27 Bachelors	3	28	14		4	6	4	8	
28 Bachelors	2	32	. 5	9	8	3 1	4	3	
29 Masters	2	29	4	10	4	8	1	8	
0 PHD	3	115	16	7	10	1	1	2	
31 Bachelors	3	18	5	10	10	1	3	2	
32 Bachelors	3	56	6	3	9	5	7	7	
33 Bachelors	3	66	13	2	. 3	3 2	7	6	

Рис. 3. Выборка из датасета II в исходном состоянии

12 3 3 3 6 6 2 8 7 3 5 13 3 3 25 6 1 5 7 4 6 14 3 3 150 15 7 9 7 7 3 15 3 1 36 6 3 8 5 6 5 16 3 3 38 12 7 1 7 6 2 18 3 3 38 12 2 5 8 1 7 18 3 2 8 4 3 7 7 9 6 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 9 8 8 7 21 4 2 3 1	Education	PaymentTier	ExperienceMonth	Average	Extroversion	Neuroticism	Kindness	Scrupulosity	Intellect	LeaveOrNot	
2 3 3 29 6 8 9 3 10 10 3 4 3 183 23 2 3 6 10 3 5 3 3 56 14 6 5 10 5 5 3 3 9 9 9 3 3 2 11 6 7 3 3 3 34 17 6 1 10 10 10 8 8 3 3 34 17 6 1 10 10 8 9 1 1 19 1 9 1 1 19 1 1 19 1 1 19 1 1 19 1 1 1 19 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1<	0	3	3	4	2	2	1	9	8	10	
3 4 3 183 23 2 3 6 10 3 4 4 3 56 14 6 6 5 10 5 5 3 3 9 9 3 3 2 11 6 8 3 3 57 7 2 2 7 7 6 8 3 3 34 17 6 1 10 10 8 9 4 2 83 21 7 4 1 1 1 9 10 4 3 180 26 8 7 2 11 3 3 1 9 1 1 1 9 1 1 1 9 1 1 1 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	3	1	78	10	3	4	5	10	9	
4 3 56 14 6 6 5 10 5 5 3 3 8 4 8 2 8 5 10 6 3 3 9 9 3 3 2 11 6 7 3 3 57 7 2 2 7 77 6 8 3 3 3 17 6 1 10 10 8 9 4 2 83 21 7 4 1 1 9 1 1 9 1 1 9 1 1 19 1 1 19 1 1 19 1 1 1 9 1	2	3	3	29	6	8	9	3	10	10	
5 3 3 9 9 9 3 2 1 6 1 6 1 6 1 6 1 6 1 1 6 1 1 6 1	3	4	3	183	23	2	3	6	10	3	
6 3 3 9 9 3 3 2 1 6 7 3 3 57 7 2 2 7 7 6 8 3 3 34 17 6 1 10 10 8 9 4 2 83 21 7 4 1 1 19 10 4 3 180 26 8 7 2 1 3 11 3 3 61 6 2 8 7 3 5 12 3 3 61 6 2 8 7 7 4 6 14 3 3 150 15 7 9 7 7 3 6 1 1 7 6 2 1 1 7 6 2 1 1 7 6 2 1 1	4	4	3	56	14	6	6	5	10	5	
7 3 3 3 34 17 6 1 10 10 8 8 3 3 34 17 6 1 10 10 8 9 4 2 83 21 7 4 1 1 9 10 4 3 180 26 8 7 2 1 3 1 3 1 3 1 3 1 3 1 3 1 3 3 1 3 3 1 4 4 4 6 4 4 6 4 4 6 4 6 4 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 6 4 7 7 7	5	3	3	8	4	8	2	8	5	10	
8 3 3 34 17 6 1 10 10 8 9 4 2 83 21 7 4 1 1 9 10 4 3 180 26 8 7 2 1 3 11 3 3 61 6 2 8 7 3 5 12 3 3 61 6 2 8 7 3 5 14 3 3 150 15 7 9 7 7 4 6 14 3 3 150 15 7 9 7 7 3 3 15 3 1 36 6 3 8 5 6 5 6 5 6 1 7 7 7 6 2 1 7 7 7 6 2 1 1 <td>6</td> <td>3</td> <td>3</td> <td></td> <td>9</td> <td>3</td> <td>3</td> <td></td> <td>1</td> <td>6</td> <td></td>	6	3	3		9	3	3		1	6	
9 4 2 83 21 7 4 1 1 9 10 4 3 180 26 8 7 2 1 3 11 3 3 82 12 10 8 9 8 3 12 3 3 61 6 2 8 7 3 5 13 3 3 150 15 7 9 7 7 3 6 14 3 3 150 15 7 9 7 7 7 3 3 16 3 8 5 6 5 6 5 6 1 7 7 7 7 6 3 6 1 7 7 7 7 6 2 1 1 7 7 6 2 1 1 8 9 9 1 6 3	7	3	3	57		2	2		7	6	
10 4 3 180 26 8 7 2 1 3 3 13 3 3 82 12 10 8 9 8 3 3 5 13 5 13 5 13 5 7 3 5 14 6 6 10 5 7 7 7 7 3 14 6 6 10 5 7 9 7 7 7 3 14 6 6 6 3 8 5 6 5 5 6 5 14 7 7 7 3 3 14 <	8	3	3	34	17	6	1	10	10	8	
11 3 3 8 12 10 8 9 8 3 12 3 3 61 6 2 8 7 3 5 13 3 3 25 6 1 5 7 4 6 14 3 3 150 15 7 9 7 7 3 15 3 1 36 6 3 8 5 6 2 17 3 3 38 12 7 6 2 18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 8 9 9 8 9 10 19 3 3 4 2 1 1 8 9 10 1 1 8 1 1 1 1 8 1	9	4	2	83	21	7	4	1	1	9	
12 3 3 3 6 6 2 8 7 3 5 13 3 3 25 6 1 5 7 4 6 14 3 3 150 15 7 9 7 7 3 15 3 1 36 6 3 8 5 6 5 16 3 3 38 12 7 1 7 6 2 18 3 3 38 12 2 5 8 1 7 18 3 2 8 4 3 7 7 9 6 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 9 8 8 7 21 4 2 3 1	10	4	3	180	26	8	7	2	1	3	
13 3 3 3 15 6 1 5 7 4 6 14 3 3 150 15 7 9 7 7 3 15 3 1 36 6 3 8 5 6 5 16 3 3 33 35 7 1 7 6 2 17 3 3 38 12 2 5 8 1 7 18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 8 9 9 8 9 20 3 2 148 21 1 9 8 8 7 21 4 2 148 21 1 9 8 8 7 22 3 3 6 2 8 9 9 10 4 5 24 3 3 6	11	3	3	82	12	10	8	9	8	3	
14 3 3 3 15 3 9 7 7 3 15 3 1 36 6 3 8 5 6 5 16 3 3 38 5 7 1 7 6 2 17 3 3 98 12 2 5 8 1 7 7 18 3 2 8 4 8 9 9 8 9 9 19 3 2 8 4 8 9 9 8 9 10 20 3 3 4 2 1 1 8 9 10 10 1 8 9 10 10 1 1 8 9 10 10 1 1 8 9 10 1 1 8 9 10 1 1 1 8 1 1 1 8 1 1 1 1 1 1 1 1	12	3	3		6	2	8	7	3	5	
15 3 1 36 6 3 8 5 6 5 16 3 3 33 5 7 1 7 6 2 17 3 3 98 12 2 5 8 1 7 18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 8 9 10 21 4 2 148 21 1 9 8 8 7 22 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 3 18 8 3 <td>13</td> <td>3</td> <td>3</td> <td></td> <td>6</td> <td>1</td> <td>5</td> <td>7</td> <td>4</td> <td>6</td> <td></td>	13	3	3		6	1	5	7	4	6	
16 3 3 3 3 5 7 1 7 6 2 17 3 3 98 12 2 5 8 1 7 18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 8 9 10 21 4 2 148 21 1 9 8 8 7 21 4 2 148 21 1 9 8 8 7 22 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 3 18 18<	14	3	3	150	15	7	9	7	7	3	
17 3 3 98 12 2 5 8 1 7 18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 8 9 10 21 4 2 148 21 1 9 8 8 7 22 3 2 81 9 3 3 2 7 6 22 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 150 25 2 2 3 8 1 1 26 3 3 28 14 8 4 6 4 8 1 27 3 3 28	15	3	1	36	6	3	8	5	6	5	
18 3 2 8 4 8 9 9 8 9 19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 8 9 10 21 4 2 148 21 1 9 8 8 7 22 3 2 81 9 3 3 2 7 6 23 3 3 150 25 2 2 3 8 1 24 3 3 150 25 2 2 3 8 1 25 3 3 154 18 8 3 9 2 5 26 3 3 2 34 8 4 6 4 8 29 4 10 4 8 1 4<	16	3	3	33	5	7	1	7	6	2	
19 3 2 8 4 3 7 7 9 6 20 3 3 4 2 1 1 8 9 10 21 4 2 11 9 8 8 7 6 22 3 2 81 9 3 3 2 7 6 23 3 3 150 25 2 2 3 8 1 24 3 3 150 25 2 2 3 8 1 25 3 3 14 4 5 4 1 3 1 26 3 3 28 14 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 29 4 10 4 8<	17	3	3	98	12	2	5	8	1	7	
20 3 3 4 2 1 1 8 9 10 21 4 2 148 21 1 9 8 8 7 22 3 2 8 9 3 2 7 6 23 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 4 4 5 4 1 3 1 26 3 3 28 14 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 3 28 14 8 4 6 4 8 29 4 2 9 4 10 4 8 1 8 31 3 3 18 5 10 10	18	3	2	8	4	8	9	9	8	9	
21 4 2 148 21 1 9 8 8 7 22 3 2 81 9 3 3 2 7 6 24 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 54 18 8 3 9 2 5 26 3 3 28 14 8 4 6 4 8 27 3 3 28 14 8 4 6 4 8 28 3 2 32 32 14 8 4 6 4 8 29 4 10 4 8 1 4 3 30 3 115 16 7 10 1 1 2 31 3 18 5 10 10 1 3 </td <td>19</td> <td>3</td> <td>2</td> <td>8</td> <td>4</td> <td>3</td> <td>7</td> <td>7</td> <td>9</td> <td>6</td> <td></td>	19	3	2	8	4	3	7	7	9	6	
22 3 2 81 9 3 3 2 7 6 23 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 4 4 5 4 1 3 1 26 3 3 54 18 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 10 4 8 1 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 18 5 10 10	20	3	3	4	2	1	1	8	9	10	
23 3 3 62 8 9 9 10 4 5 24 3 3 150 25 2 2 3 8 1 25 3 3 4 4 5 4 1 3 1 26 3 3 54 18 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 23 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 6 6 3 <td< td=""><td>21</td><td>4</td><td>2</td><td>148</td><td>21</td><td>1</td><td>9</td><td>8</td><td>8</td><td>7</td><td></td></td<>	21	4	2	148	21	1	9	8	8	7	
24 3 3 150 25 2 2 3 8 1 25 3 3 4 4 5 4 1 3 1 26 3 3 54 18 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 5 6 3 9 5 7 7 7	22	3	2	81	9	3	3	2	7	6	
25 3 3 4 4 5 4 1 3 1 26 3 3 54 18 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 5 6 3 9 5 7 7	23	3	3	62	8	9	9	10	4	5	
26 3 3 54 18 8 3 9 2 5 27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7	24	3	3	150	25	2	2	3	8	1	
27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7	25	3	3	4	4	5	4	1	3	1	
27 3 3 28 14 8 4 6 4 8 28 3 2 32 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7	26	3	3	54	18	8	3	9	2	5	
28 3 2 32 5 9 8 1 4 3 29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 5 6 3 9 5 7 7		3				8	4		4		
29 4 2 29 4 10 4 8 1 8 30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7							8		4		
30 5 3 115 16 7 10 1 1 2 31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7									1		
31 3 3 18 5 10 10 1 3 2 32 3 3 56 6 3 9 5 7 7									1		
32 3 3 56 6 3 9 5 7 7											
					-				-		
	33	3	3	66	13	2	3	2	7	6	

Рис. 4. Выборка из датасета II после нормализации и балансировки

Таблица 2 Критерии соответствия кандидата должности и формулы для их определения

Критерий	Обозначение	Формула
Уровень образования / Ученая степень	$R_{ m edu}$	$\frac{\text{edu}_{\text{user}}}{\text{edu}_{\text{vacancy}}} \text{sign}_{\text{edu}} \cdot 10,$
		где edu _{user} – уровень образования соискателя; edu _{vacancy} – требуемый уровень образования для трудоустройства на должность; sign _{edu} – уровень значимости критерия «Уровень образования» (указывается работодателем)

Окончание табл. 2

Критерий	Обозначение	Окончание табл. 2 Формула
Опыт работы	R _{exp}	• • • • • • • • • • • • • • • • • • • •
•	Cap	$\frac{2 \exp_{\text{user}}}{\exp_{\text{vacancy-min}} + \exp_{\text{vacancy-max}}} \text{sign}_{\exp} \cdot 10,$
		при $\exp_{\text{user}} \le \frac{\exp_{\text{vacancy-min}} + \exp_{\text{vacancy-max}}}{2}$
		или
		$\frac{\exp_{\text{vacancy-min}} + \exp_{\text{vacancy-max}}}{2 \exp_{\text{user}}} \operatorname{sign}_{\exp} \cdot 10,$
		при $\exp_{user} > \frac{\exp_{vacancy-min} + \exp_{vacancy-max}}{2}$,
		где ехр _{иser} — опыт работы соискателя; ехр _{vacancy-min} — минимальный требуемый опыт работы для трудоустройства на должность; ехр _{vacancy-max} — максимальный требуемый опыт работы для трудоустройства на должность;
		sign _{exp} – уровень значимости критерия «Опыт работы» (указывается работодателем)
Зарплатные ожидания	$R_{ m sal}$	$\frac{2 \operatorname{sal}_{\operatorname{user}}}{\operatorname{sal}_{\operatorname{vacancy-min}} + \operatorname{sal}_{\operatorname{vacancy-max}}} \operatorname{sign}_{\operatorname{sal}} \cdot 10,$
		при $\exp_{user} \leq \frac{sal_{vacancy-min} + sal_{vacancy-max}}{2}$ или
		$\frac{\mathrm{sal}_{\mathrm{vacancy-min}} + \mathrm{sal}_{\mathrm{vacancy-max}}}{2\mathrm{sal}_{\mathrm{user}}}\mathrm{sign}_{\mathrm{sal}} \cdot 10,$
		при $\operatorname{sal}_{\operatorname{user}} > \frac{\operatorname{sal}_{\operatorname{vacancy-min}} + \operatorname{sal}_{\operatorname{vacancy-max}}}{2}$,
		где sal _{user} – зарплатные ожидания соискателя; sal _{vacancy-min} – минимальный размер зарплаты, предлагаемый работодателем;
		sal _{vacancy-max} — максимальный размер зарплаты, предлагаемый работо- дателем; sign _{sal} — уровень значимости критерия «Зарплатные ожидания»
Формат работы	R_{F}	(указывается работодателем) (cor _{form} + cor _{schedule}) sign _F · 10,
		где сог _{form} – корреляция по форме занятости (0 или 1); сог _{schedule} – корреляция по графику (0 или 1); sign _F – уровень значимости критерия «Формат работы» (указывается работодателем)
Мобильность (воз- можность переезда	$R_{ m M}$	$(cor_{duty} + cor_{relocation}) sign_M \cdot 10,$
и командировок)		где сог _{duty} – корреляция по возможности командировок (0 или 1); сог _{relocation} – корреляция по возможности переезда (0 или 1); sign _M – уровень значимости критерия «Мобильность» (указывается работодателем)
Рейтинг соискателя	R _{Total}	$R_{\text{edu}} + R_{\text{exp}} + R_{\text{sal}} + R_{\text{F}} + R_{\text{M}} + \text{CosSim},$
		где CosSim – косинусное сходство между требуемыми навыками и навыками соискателя (от 0,00 до 10,00)

В целях демонстрации возможностей системы было создано 4 набора данных, каждый из которых содержит сведения о 200 кандидатах.

Таким образом, в систему было загружено 800 записей о тестовых кандидатах, содержащие сведения об уровне образования, опыте работы,

надежности (средняя продолжительность трудовых отношений с одним работодателем), зарплатных ожиданиях, готовности к переезду и командировкам, предпочитаемых формах занятости и графике работы. Также созданные наборы данных включали информацию о навыках кандидатов и результатах тестирования личности в соответствии с методикой BigFive. При использовании математических методов вычислялся рейтинг кандидатов (рейтинг МАИ), а при использовании моделей случайного леса прогнозировались целесообразность рассмотрения кандидата на должность (датасет I лежит в основе показателя «Вывод ИИ») и ожидаемая длительность трудовых отношений (датасет II лежит в основе показателя «Ожидаемая длительность трудовых отношений»). Сведения представлены на рис. 6 и 7.

ВЫВОДЫ

Предложена методика анализа соответствия соискателя рассматриваемой должности, в основе которой лежит использование математических методов и обученных моделей случайного леса.

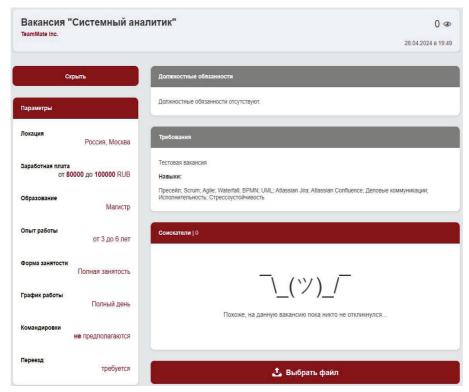


Рис. 5. Требования должности: базовые параметры и компетенции



Рис. 6. Оценка кандидатов, рекомендуемых системой к рассмотрению



Рис. 7. Оценка кандидатов, не рекомендуемых системой к дальнейшему рассмотрению

СПИСОК ИСТОЧНИКОВ

- 1. Клеткина Н. В. Актуальность и методы отбора персонала. Проблемы и пути их решения // Научный журнал. 2017. URL: https://cyberleninka.ru/article/n/aktualnost-i-metody-otbora-personala-problemy-i-puti-ih-resheniya (дата обращения: 10.04.2024).
- 2. Белова Е. Анализируем резюме // HR-академия. URL: https://hr-academy.ru/hrarticle/analiziruem-rezyume.html (дата обращения: 10.04.2024).
- 3. Стародубцева М. Анализ резюме: как рекрутеру понять, стоит ли приглашать кандидата на собеседование // PAMPADU : онлайн-платформа. URL: https://pampadu.ru/blog/4707-analiz-rezyume-kak-rekruteru-ponyat-stoit-li-priglashat-kandidata-na-sobesedovanie (дата обращения: 10.04.2024).
- 4. Профессиональный анализ резюме // Клерк.ру. URL: https://www.klerk.ru/blogs/rosco/501251 (дата обращения: 10.04.2024).
- 5. Analytic hierarchy process // Wikipedia. URL: https://en.wikipedia.org/wiki/Analytic_hierarchy_process (дата обращения: 10.04.2024).
- 6. Cosine similarity // Wikipedia. URL: https://en.wikipedia.org/wiki/Cosine_similarity (дата обращения: 10.04.2024).

REFERENCES

- 1. Kletkina N. V. Actuality and methods of employee recruitment. Problems and ways of their solution. *Nauchniy zhurnal* [Scientific journal]. 2017. URL: https://cyberleninka.ru/article/n/aktualnost-i-metody-otbora-personala-problemy-i-puti-ih-resheniya (accessed 10.04.2024). (In Russ.)
- 2. Belova E. Analyzing a resume. URL: https://hr-academy.ru/hrarticle/analiziruem-rezyume.html (accessed 10.04.2024). (In Russ.)
- 3. Starodubceva M. Resume analysis: how can a recruiter understand whether it is worth inviting a candidate for an interview. URL: https://pampadu.ru/blog/4707-analiz-rezyume-kak-rekruteru-ponyat-stoit-li-priglashat-kandidata-na-sobesedovanie (accessed 10.04.2024). (In Russ.)
- 4. Professional resume analysis. URL: https://www.klerk.ru/blogs/rosco/501251 (accessed 10.04.2024). (In Russ.)
- 5. Analytic hierarchy process. URL: https://en.wikipedia.org/wiki/Analytic_hierarchy_process (accessed 10.04.2024).
- 6. Cosine similarity URL: https://en.wikipedia.org/wiki/Cosine similarity (accessed 10.04.2024).

Статья поступила в редакцию 16.05.2024 Принята к публикации 24.05.2024