

УДК 004.023

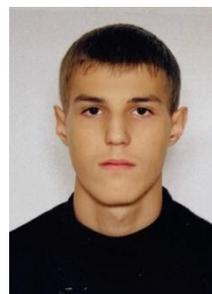
РАЗРАБОТКА НЕЙРОСЕТЕВОГО МОДУЛЯ ПРОГНОЗИРОВАНИЯ СТРАХОВЫХ ВЫПЛАТ



Т.А. Васяева
Декан факультета
информационных систем и
технологий, доцент
кафедры
автоматизированных
систем управления
ДОННТУ,
кандидат технических наук,
доцент
vasyaeva@gmail.com



Ю.А. Золушкин
Аспирант кафедры
автоматизированных
систем управления
ДОННТУ
illuzium1999@gmail.com



А.В. Яковчук
Магистрант кафедры
автоматизированных
систем управления
ДОННТУ
Yakovchuk-2001@mail.ru

Т.А. Васяева

Окончила Донецкий национальный технический университет. Область научных интересов связана с разработкой методов и алгоритмов искусственного интеллекта, организацией учебного и научно-исследовательского процессов в техническом университете.

Ю.А. Золушкин

Окончил Донецкий национальный технический университет. Область научных интересов связана с разработкой методов машинного обучения, в том числе алгоритмов обработки естественного языка.

А.В. Яковчук

Магистрант Донецкого национального технического университета. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, машинного обучения.

Аннотация. Решается задача повышения точности планирования финансирования деятельности страхового Фонда, за счет прогнозирования его расходов. Предложено использовать нейросетевой подход для прогнозирования расходов Фонда. Выполнен предварительный анализ имеющихся статистических данных. Предложена нейросетевая модель для прогнозирования страховых выплат. Проведены эксперименты по подбору гиперпараметров нейронной сети.

Ключевые слова: нейронная сеть, временной ряд, автокорреляционная функция, социальное страхование.

Введение. Социальное страхование сегодня, в современном обществе, играет большую роль в функционировании государства и экономики, а также в поддержке жизненного уровня населения. Социальное страхование является одним из институтов экономического развития общества, независимо от его политического устройства и статуса на мировой арене. Основа развития страхового рынка заключается в необходимости обеспечения бесперебойности воспроизведенного процесса путем

оказания денежной помощи пострадавшим в случае непредвиденных неблагоприятных обстоятельств.

Обязательное условие страхового рынка - наличие общественной потребности на страховые услуги и наличие страховщиков, способных удовлетворить эти потребности.

В период не стабильного развития экономики часто возникают ситуации, когда происходит неправильное и не точное планирования бюджетных средств, что влечет за собой нехватку финансирования для выполнения обязательств. Для того, чтобы предугадать поведение внутри финансирования страхового Фонда осуществляют прогнозирование финансовых показателей [1].

Анализ предметной области и цель исследования. Фонд социального страхования от несчастных случаев на производстве и профессиональных заболеваний Донецкой Народной Республики (далее – Фонд) является некоммерческой самоуправляющейся организацией, которая осуществляет профилактическую работу по обеспечению безопасных условий труда на предприятиях и учреждениях республики и управляет средствами государственного социального страхования от несчастных случаев на производстве и профессиональных заболеваний.

Для исследования было выбрано отделение Фонда социального страхования от несчастных случаев на производстве и профессиональных заболеваний Донецкой Народной Республики в Кировском районе г. Макеевки.

Отделение Фонда обязано произвести прогнозирование и планирование основных расходов на месяц, квартал, год и передать данные в Центральный аппарат. Центральный аппарат Фонда в свою очередь, выделяет бюджетные средства для функционирования деятельности отделения, которое в дальнейшем производит выплаты и закупку средств медицинской реабилитации потерпевшим, которые состоят на учете в Фонде.

Основными видами расходов отделения Фонда являются:

- страховая выплата в случае частичной или полной утраты трудоспособности;
- расходы на лекарственные средства и изделия медицинского назначения;
- расходы на поддержание существования Фонда (аренда, коммунальные услуги, зарплата и т.д.).

От правильности и точности прогнозирования и планирования зависит объем финансирования, которое выделит Центральный аппарат, поэтому погрешность должна быть минимальной.

Целью исследования является повышение точности планирования финансирования деятельности страхового Фонда, за счет прогнозирования его расходов.

Постановка задачи и данные для исследования. В качестве исследуемых данных в условиях отделения Фонда были собраны статистические данные за 3 года, которые представляют собой суммы ежемесячных страховых выплат; ежемесячных выплат по медицине; ежемесячных расходов на содержание Фонда в российских рублях. Собранные параметры представляют собой одномерные временные ряды с интервалом в один месяц.

Наша задача сводится к прогнозированию одномерных временных рядов. Задача прогнозирования временных рядов заключается в следующем. Заданы дискретные отсчеты $\{y(t_1), y(t_2), \dots, y(t_n)\}$ в последовательные моменты времени t_1, t_2, \dots, t_n . Необходимо рассчитывать значения $y(t_{n+1})$ в некоторый будущий момент времени t_{n+1} .

Анализ данных. При анализе и изучении временного ряда аналитик должен на основе некоторого отрезка ряда конечной длины сделать выводы о характере и закономерностях процесса, который описывается данным рядом. Таким образом, цель анализа временного ряда – описание характерных особенностей ряда, для дальнейшего построения модели ряда для предсказания будущих значений на основе прошлых наблюдений [2].

Для этого будем использовать аппарат корреляционного анализа. Корреляция – взаимосвязь двух или более величин, при этом изменения значений одной из величин сопровождается изменению значений другой величины. При анализе временных рядов вычисляется автокорреляция ряда и строится автокорреляционная функция [3].

Таблица 1. Данные для расчета автокорреляционной функции (АКФ) на примере данных по ежемесячным выплатам по медицине

X	355775	361541	352862	342719	...	339457	363758	376256
X_{i-1}		355775	361541	352862	...	350852	339457	363758
X_{i-2}			355775	361541	...	331852	350852	339457
X_{i-3}				355775	...	328785	331852	350852
X_{i-4}					...	356581	328785	331852
...
X_{i-n}					...			355775

Коэффициент автокорреляции изменяется в диапазоне $[-1;1]$, где $r_k = 1$ означает полную корреляцию. Рассчитав коэффициенты автокорреляции для каждого сдвига, получим автокорреляционную функцию (АКФ). Автокорреляционная функция – последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков. Расчет АКФ на примере данных по ежемесячным выплатам по медицине, для ежемесячных страховых выплат и ежемесячных расходов на содержание Фонда в табл. 2-4 и графики АКФ на рис. 1.

Таблица 2. Результаты АКФ по ежемесячным страховым выплатам.

k	0	1	2	3	4	5	6	7	8	9	10	11
r_k	1	0,99	0,98	0,98	0,98	0,97	0,97	0,97	0,96	0,97	0,973	0,98
k	12	13	14	15	16	17	18	19	20	21	22	23
r_k	0,98	0,98	0,98	0,98	0,97	0,95	0,93	0,921	0,914	0,91	0,90	0,90
k	24	25	26	27	28	29	30	31	32	33	34	35
r_k	0,92	0,92	0,93	0,94	0,94	0,92	0,84	0,761	0,72	0,67	0,62	0,67

Таблица 3. Результаты АКФ по выплатам по медицине

k	0	1	2	3	4	5	6	7	8	9	10	11
r_k	1	0,07	0,07	0,5	-0,3	-0,5	0,3	-0,6	-0,3	0,4	-0,1	-0,00
k	12	13	14	15	16	17	18	19	20	21	22	23
r_k	0,92	0,07	0,12	0,52	-0,26	-0,5	0,26	-0,64	-0,34	0,35	-0,15	-0,05
k	24	25	26	27	28	29	30	31	32	33	34	35
r_k	0,89	0,02	0,13	0,6	-0,27	-0,45	0,28	-0,7	-0,46	0,36	-0,27	-0,13

Таблица 4. Результаты АКФ по расходам на содержание Фонда

k	0	1	2	3	4	5	6	7	8	9	10	11
r_k	1	-0,04	0,15	0,03	-0,19	-0,34	0,02	-0,41	-0,28	0,098	0,056	0,07
k	12	13	14	15	16	17	18	19	20	21	22	23
r_k	0,78	-0,1	0,08	0,16	-0,13	-0,15	-0,05	-0,48	-0,22	-0,02	0,04	0,14
k	24	25	26	27	28	29	30	31	32	33	34	35
r_k	0,72	-0,09	0,35	0,19	-0,09	-0,18	-0,24	-0,54	-0,19	-0,12	0,06	0,07

Проанализировав полученные результаты, можно сделать вывод, что временной ряд по ежемесячным страховым выплатам содержат тренд, так как степень взаимной зависимости между последовательными значениями ряда и корреляция между ними очень высоки. При этом коэффициент автокорреляции значителен для многих сдвигов ряда.

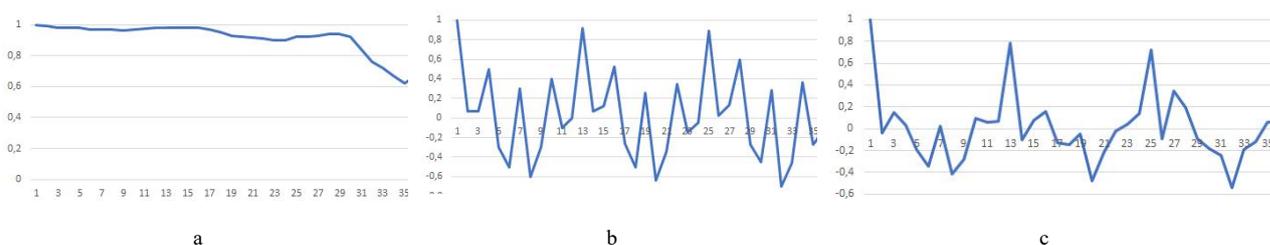


Рисунок 1. Автокорреляционная функция временного ряда *a*) по ежемесячным страховым выплатам, *b*) по выплатам по медицине, по расходам на содержание Фонда

Скорее всего, временной ряд по выплатам по медицине и по расходам на содержание Фонда содержат сезонную компоненту, так как коэффициент автокорреляции периодически увеличивается. Наличие сезонной компоненты можно предположить и по визуальному анализу графиков временных рядов, но так как данные всего за три года, а предположительный период равен году, то недостаточно данных для однозначного утверждения наличия сезонной компоненты.

Учитывая наличие тренда и высоких значений коэффициентов автокорреляции для многих сдвигов ряда, можно предположить, что для построения прогностической модели можно брать порядка 10 предшествующих значений для временного ряда по ежемесячным страховым выплатам. Но учитывая ограниченное количество данных, остановимся на 5 входах, так как увеличение числа входов существенно сократит обучающую выборку. По выплатам по медицине и по расходам на содержание Фонда будем так же использовать 5 входов.

Обобщенная модель и построение обучающей выборки. Обобщенная модель прогнозирования временного ряда, имеет следующий вид (рис. 2).

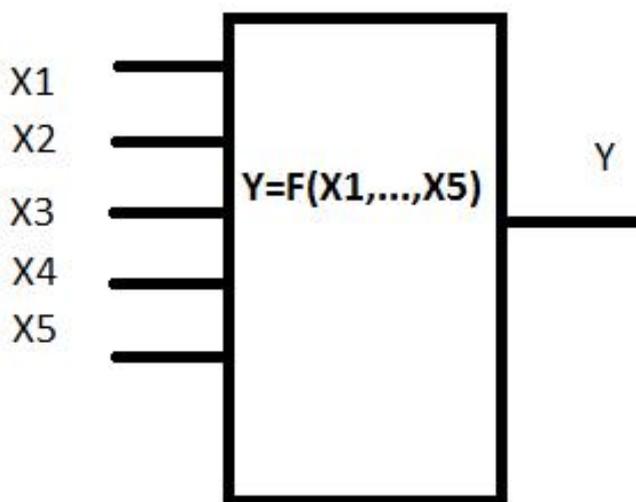


Рисунок 2. Обобщенная модель для прогнозирования временного ряда по ежемесячным страховым выплатам

Обучающая выборка строится методом скользящего окна [3], который широко применяется при работе с временными рядами. Обучающие выборки для такой модели будут иметь такой формат (табл. 5).

Разработка нейросетевой модели и ее оптимизация. Для прогнозирования временных рядов, как показывает опыт, хорошо показывают себя многослойные

нейронные сети. Можно также предположить, что 3-х слоев будет достаточно, затем проверить и подтвердить экспериментально. На выходном слое будет один нейрон, так как прогнозируем на один шаг (месяц) вперед. Количество нейронов на скрытых слоях будем определять экспериментально. К другим гиперпараметрам модели относят: количество эпох обучения, функции активации нейронов, функция ошибок. Так же будем экспериментально подбирать оптимизатор.

Таблица 5. Фрагмент обучающей выборки по ежемесячным страховым выплатам

X_1	X_2	X_3	X_4	X_5	Y
1901542	1913564	1913564	1926764	1982457	1984254
1913564	1913564	1926764	1982457	1984254	1995424
1913564	1926764	1982457	1984254	1995424	1997403
...
2265412	2267521	2300452	2357521	2378552	2378552

В среде *Colab Laboratory* выполнена программная реализация нейронной сети для прогнозирования. Для моделирования использовался язык *Python*, с применением библиотек *pandas*; *numpy*; *keras* (в качестве основы *tensorflow*).

Colab Laboratory [5] – это бесплатная среда для ноутбуков *Jupyter*, которая не требует настройки и полностью работает в облаке. *Colab Laboratory* позволяет писать и выполнять код, а также получать доступ к мощным вычислительным ресурсам, что является важным преимуществом при работе с ресурсоемкими алгоритмами машинного обучения.

Keras [6] – одна из самых мощных и простых в использовании библиотек *Python* для разработки и оценки моделей глубокого обучения, охватывает эффективные библиотеки численных вычислений *Theano* и *TensorFlow*. Преимущество этого состоит главным образом в том, есть возможность работать с нейронными сетями достаточно простым способом.

Экспериментальные исследования. Проведем эксперименты по выбору архитектуры НС и настройки основных гиперпараметров модели.

Для начала проведем исследования для данных ежемесячных страховых выплат. Структура сети будет следующая: 5 входов (согласно данным), 5 нейронов на первом слое, 6 нейронов на втором слое и 1 нейрон на выходном слое, функции активации: *relu*, *relu*, *linear*. В таблице 6 представлены эксперименты на обучающей, валидационной, тестовой выборках при разном количестве эпох. Точность анализируем по *Mean Squared Error (MSE)* и *Mean absolute error (MAE)*. *MSE* рассчитывается как среднее значение квадратов разностей между прогнозируемыми и фактически наблюдаемыми значениями. *MAE* рассчитывается как среднее значение абсолютной разницы между прогнозами и фактическими наблюдениями. Оптимизатор будем использовать *adam*. Результаты экспериментов представлены в таблице 6. Как видно из таблицы 6, на 500 эпохе наблюдаем переобучение сети, следовательно дальнейшее обучение нецелесообразно.

Далее, используя ту же архитектуру, проведем эксперименты по ежемесячным выплатам по медицине. Результаты экспериментов представлены в таблице 7. Анализируя таблицу 7, можно заметить, что при модели обучается плохо. Уже на 100 эпохе появляются признаки переобучения, далее модель обучается незначительно, а при обучении на 500 эпохе опять наблюдаем признаки переобучения.

Таким же образом, проведем исследования на данных по ежемесячным расходам на содержание Фонда. Результаты экспериментов представлены в таблице 8. Сеть практически не обучается.

Изменим архитектуру нейронной сети на следующую: 5 входов, 5 нейронов на

первом слое, 6 нейрона на втором слое, 1 на выходном; функции активации: гиперболический тангенс, sigmoid, relu. Результаты экспериментов представлены в таблицах 9-11.

Таблица 6 Ошибка обучения на НС по ежемесячным страховым выплатам

epochs	Training set		Validation set		Test set	
	mse	mae	mse	mae	mse	mae
25	0.0194	0.1099	0.2645	0.5118	0.0053	0.6201
50	0.0167	0.1029	0.2253	0.4713	0.5226	0.7126
75	0.0032	0.0488	0.0292	0.1675	0.4930	0.6937
100	0.0012	0.0290	0.0050	0.0628	0.0607	0.2383
200	0.0016	0.0307	0.0049	0.0543	0.0060	0.0658
300	0.0007	0.0192	0.0014	0.0292	0.0060	0.0603
400	0.0013	0.0280	0.0010	0.0296	0.0014	0.0359
500	0.0008	0.0192	0.0033	0.0511	0.0042	0.0557

Таблица 7 Ошибка обучения на НС по ежемесячным выплатам по медицине

epochs	Training set		Validation set		Test set	
	mse	mae	mse	mae	mse	mae
25	0.1972	0.3793	0.2299	0.4109	0.2058	0.3852
50	0.1331	0.3138	0.1889	0.3851	0.2979	0.4668
75	0.0232	0.1262	0.0267	0.0906	0.2057	0.3875
100	0.0577	0.2014	0.0754	0.2437	0.0509	0.1988
200	0.0235	0.1226	0.0277	0.1164	0.0917	0.2566
300	0.0200	0.1161	0.0241	0.1207	0.0511	0.2190
400	0.0187	0.1192	0.0331	0.1341	0.0454	0.1878
500	0.0544	0.1887	0.0838	0.2270	0.0429	0.1920

Таблица 8 Ошибка обучения на НС по ежемесячным расходам на содержание Фонда

epochs	Training set		Validation set		Test set	
	mse	mae	mse	mae	mse	mae
25	0.1038	0.2661	0.2627	0.4759	0.0426	0.1632
50	0.1273	0.2808	0.1323	0.3163	0.0358	0.1597
75	0.0551	0.1954	0.1352	0.3215	0.0620	0.2016
100	0.0607	0.2085	0.0904	0.2693	0.0404	0.1739
200	0.0592	0.1913	0.0807	0.2215	0.0429	0.1668
300	0.0507	0.1945	0.0946	0.2600	0.0478	0.1871
400	0.0483	0.1855	0.0982	0.2451	0.0536	0.1901
500	0.0223	0.1014	0.1322	0.3145	0.0543	0.1845

Как видно из таблицы 9, на 500 эпохе наблюдаем переобучение сети, следовательно дальнейшее обучение нецелесообразно. Так же отмечаем, что нелинейные функции активации на скрытых слоях дают лучшие результаты. Изменим оптимизатор на *adamax* и проведем эксперименты (табл. 10.), а затем применим оптимизатор *SGD* (табл. 11).

Анализируя данные, можно сказать, что оптимизатор *SGD* показывает лучшие результаты. При этом следует отметить, что сеть имеет склонность к переобучению.

Аналогичным образом был выполнен еще ряд экспериментов. Отмечено, что увеличение количества нейронов и / или количество слоев приводит к переобучению при меньшем количестве эпох, но не способствуют улучшению результата (снижению ошибки).

Таблица 9 Ошибка обучения на НС оптимизатор *adam*

<i>epochs</i>	<i>Training set</i>		<i>Validation set</i>		<i>Test set</i>	
	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>
25	0.0420	0.0236	0.0203	0.0134	0.0094	0.3545
50	0.0210	0.0228	0.0148	0.0148	0.1430	0.1213
75	0.0530	0.0537	0.03523	0.1571	0.1967	0.1039
100	0.0160	0.0259	0.03523	0.0852	0.1967	0.2648
200	0.0148	0.0176	0.1375	0.0645	0.1967	0.2849
300	0.0170	0.0196	0.0714	0.0215	0.0194	0.2460
400	0.0162	0.0147	0.0950	0.0148	0.1967	0.1319
500	0.0168	0.0139	0.0600	0.0148	0.1967	0.1039

Таблица 10 Ошибка обучения на НС оптимизатор *adamax*

<i>epochs</i>	<i>Training set</i>		<i>Validation set</i>		<i>Test set</i>	
	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>
25	6.9462e-05	0.0135	0.1039	0.2313	0.1967	0.4219
50	0.0002	0.0135	0.0497	0.0572	0.0045	0.1177
75	6.7806e-05	0.0238	0.1039	0.2313	0.1967	0.1180
100	3.5834e-05	0.0120	0.1039	0.2313	0.1967	0.1185
200	0.0006	0.0057	0.1039	0.2313	0.1967	0.0755
300	1.9409e-05	0.0232	0.1039	0.2313	0.1967	0.3214
400	6.1079e-05	0.0032	0.1039	0.2313	0.1967	0.4219
500	0.0002	0.0025	0.0331	0.0062	0.0012	0.4219

Таблица 11 Ошибка обучения на НС оптимизатор *SGD*

<i>epochs</i>	<i>Training set</i>		<i>Validation set</i>		<i>Test set</i>	
	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>	<i>mse</i>	<i>mae</i>
25	0.0006	0.0020	0.1101	0.0148	0.1751	0.0208
50	0.0032	0.0019	0.0538	0.0148	0.1297	0.0210
75	0.0004	0.0023	0.0538	0.0032	0.0261	0.0211
100	0.0004	0.0023	0.0538	0.0148	0.0148	0.1967
200	0.0005	0.0017	0.0038	0.0148	0.1571	0.0205
300	0.0004	0.0016	0.0538	0.0148	0.0852	0.0205
400	0.0005	0.0015	0.0538	0.0148	0.0645	0.0210
500	0.0005	0.0022	0.0538	0.0148	0.0215	0.0094

Заключение. В данной работе предложен подход к прогнозированию основных видов расходов отделения Фонда. Предложено использовать нейросетевую модель для прогнозирования временных рядов страховых выплат. Проведены эксперименты по разработке нейронной сети согласно которым можно сделать вывод: рекомендовано использовать нелинейные функции активации на скрытых слоях и *relu* на выходном слое; оптимизатор *SGD*; достаточно использовать 3 слоя нейронной сети. Разработана архитектура НС: 5 входов, 5 нейронов на первом слое, 6 нейрона на втором слое, 1 на выходном; функции активации: гиперболический тангенс, *sigmoid*, *relu*, оптимизатор *SGD*.

Список литературы

[1]. Васяева Т.А., Шуватова Е.А. Яковчук А.В. Анализ методов прогнозирования для управления финансовыми средствами страховой компании. Информатика, управляющие системы, математическое и компьютерное моделирование» (ИУСМКМ-2023): сборник трудов XIV международной научно-технической

конференции в рамках IX Международного Научного форума Донецкой Народной Республики / Под ред. Р.В. Мальчевой, И.В. Матях; ФГБОУ ВО «ДонНТУ». – Донецк: ДонНТУ, 2023.

[2]. Peter J. Brockwell, Richard A. Davis: Introduction to Time Series and Forecasting. Second Edition. Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA (2002).

[3]. Sergii, K., Yurii, S., Tatyana, V., Natalia, A.: Feature Selection for Time-Series Prediction in Case of Undetermined Estimation. In: Samsonovich A., Klimov V., Rybina G. (eds.) Biologically Inspired Cognitive Architectures (BICA) for Young Scientists, Advances in Intelligent Systems and Computing, vol. 449, pp. 85-97. Springer, Cham (2016). DOI: 10.1007/978-3-319-32554-5_12

[4]. Vasyaeva, T., Martynenko, T., Khmilovyi, S., Andrievskaya, N. (2019). Stock Prices Forecasting with LSTM Networks. In: Kuznetsov, S., Panov, A. (eds) Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science, vol 1093. Springer, Cham. https://doi.org/10.1007/978-3-030-30763-9_5

[5]. Google Colaboratory. - <https://colab.research.google.com/notebooks/welcome.ipynb>.

[6]. Keras: The Python Deep Learning library. - <https://keras.io>.

Авторский вклад

Васяева Татьяна Александровна – руководство исследованием по прогнозированию основных видов расходов отделения Фонда социального страхования.

Золушкин Юрий Алексеевич – реализация алгоритма и разработка нейросетевой модели прогнозирования временных рядов.

Яковчук Андрей Владимирович – тестирование нейросетевого модуля прогнозирования страховых выплат.

DEVELOPMENT OF A NEURAL NETWORK MODULE FOR PREDICTING INSURANCE PAYMENTS

T.A. Vasiaeva

Dean of the Faculty of Information Systems and Technologies, Associate Professor of the Department of Automated Control Systems of DONNTU, PhD of Technical Sciences, Associate Professor

Yu.A. Zolushkin

Postgraduate student of the department automated control systems of DONNTU

A.V. Yakovchuk

Master's student of the department automated control systems of DONNTU

Abstract. The problem of improving the accuracy of planning the funding of the activities of the Insurance Fund by forecasting its expenses is being addressed. A neural network approach is proposed for predicting the Fund's expenses. A preliminary analysis of the existing statistical data has been carried out. A neural network model for predicting insurance payouts has been proposed. Experiments have been conducted to tune the hyperparameters of the neural network.

Keywords: neural network, time series, autocorrelation function, social insurance.