

Исследование применимости линейной регрессии для построения прогноза расходов технико-экономических показателей котельных

Михайлова Екатерина Олеговна

Владимирский государственный университет им. А. Г. и Н. Г. Столетовых
Студент

Васильев Дмитрий Николаевич

Владимирский государственный университет им. А. Г. и Н. Г. Столетовых
Доцент

Аннотация

В статье рассматриваются этапы построения модели прогнозирования. Изучается применение созданной модели прогнозирования для расчета технико-экономических показателей котельной с помощью линейной регрессии.

Ключевые слова: прогноз, линейная регрессия, модель.

The study of the applicability of linear regression to forecast the cost of technical and economic indicators of boilers

Mikhailova Ekaterina Olegovna

Vladimir State University A. G. and N. G. Stoletovs
Student

Vasiliev Dmitry Nikolaevich

Vladimir State University A. G. and N. G. Stoletovs
Associate professor

Abstract

The article discusses the stages of building a forecasting model. Application of the created model of forecasting for calculation of technical and economic indicators of a boiler room by means of linear regression is studied.

Keywords: forecast, linear regression, model.

Прогнозирование на сегодняшний момент является одной из самых востребованных задач в самых разнообразных областях человеческой деятельности.

Развитие методов прогнозирования напрямую связано с развитием информационных технологий, в частности, с ростом объема хранимых данных и усовершенствованием методов и алгоритмов прогнозирования, реализованных в различных инструментах Data Mining.

Прогнозирование направлено на определение тенденций динамики конкретного объекта или события на основе ретроспективных данных, то есть анализа его состояния в прошлом и настоящем времени. Таким образом, решение задачи прогнозирования требует обучающей выборки данных.

На практике очень сложно получить качественный прогноз за один шаг. Технология построения прогноза включает в себя несколько действий: 1) очистка данных от помех и искажений; 2) трансформация данных; 3) построение модели; 4) решение задачи прогнозирования на построенной модели для заданного горизонта прогнозирования; 5) интерпретация результатов полученного прогноза.

Задача статьи заключается в исследовании применимости линейной регрессии для прогнозирования технико-экономических показателей котельных (ТЭПК).

Для качественного прогноза необходимо обрабатывать большие массивы данных. Поэтому для уменьшения человеческого фактора и получения более релевантного прогноза ресурсов ТЭПК необходимо использовать специализированные программные решения, в качестве которого выбрана аналитическая платформа Deductor.

Входные данные для прогнозирования - данные о расходах ТЭПК прошлых периодов. Исследование применимости метода предполагает сравнение фактических данных за текущий год с полученными прогнозными значениями.

Первым шагом в реализации задачи выполняется подготовка данных для дальнейшей загрузки их в Deductor. В таблице 1 представлена форма, в которой исходные сведения одного из показателей ТЭПК за несколько лет импортируются в аналитическую систему.

Таблица 1 – Входные данные

Дата	Объем	Показатель
01.01.2014	690,364	Ресурсы тепловой энергии, Гкал
01.02.2014	506,2	Ресурсы тепловой энергии, Гкал
....
01.11.2018	462,591	Ресурсы тепловой энергии, Гкал
01.12.2018	505,259	Ресурсы тепловой энергии, Гкал

На рисунке 1 представлен импорт подготовленных данных по расходам ресурса «Ресурсы тепловой энергии, Гкал» в Deductor.

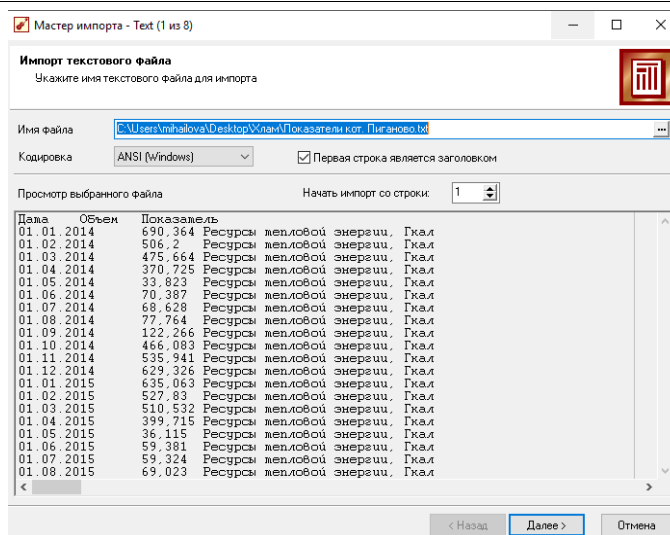


Рисунок 1 - Импорт исходных данных

Следующий шаг - использование мастера обработки «Настройка набора данных». Настройка полей данных позволяет изменять имена, метки, типы и назначения полей исходной выборки данных. В качестве способа отображения загруженных данных выбирается OLAP куб и по нему строится кросс-диаграмма (рисунок 2).

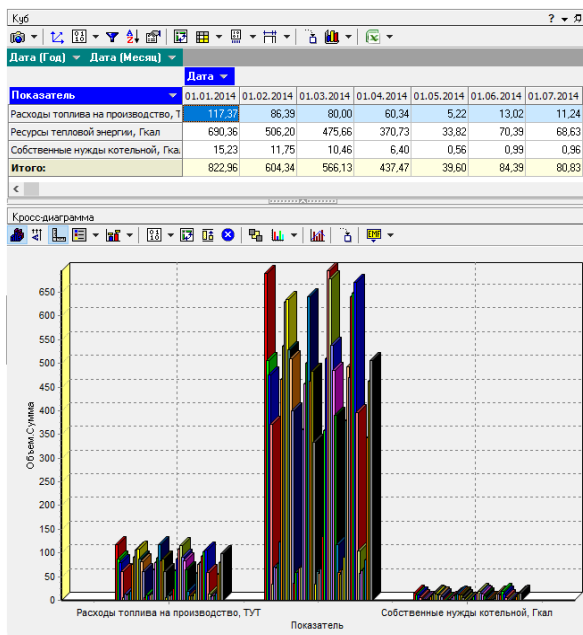


Рисунок 2 - Динамика расходов ресурсов

По диаграмме невозможно сделать вывод о сезонности расходов ТЭПК, поэтому для подтверждения или опровержения предположения о наличие сезонности в исходных данных используется автокорреляция.

Перед выполнением автокорреляции выбираются данные по одному показателю из всего набора. Для этого в мастере обработки выбирается пункт «фильтрация данных» и определяется условие «Показатель = Расходы топлива на производство, ТУТ» после чего данные фильтруются по условию.

01.09.2016	Расходы топлива на производство, ТУТ	24,551
01.10.2016	Расходы топлива на производство, ТУТ	63,112
01.11.2016	Расходы топлива на производство, ТУТ	86,953
01.12.2016	Расходы топлива на производство, ТУТ	108,145
01.01.2017	Расходы топлива на производство, ТУТ	115,842
01.02.2017	Расходы топлива на производство, ТУТ	91,463
01.03.2017	Расходы топлива на производство, ТУТ	82,925
01.04.2017	Расходы топлива на производство, ТУТ	64,2
01.05.2017	Расходы топлива на производство, ТУТ	18,347
01.06.2017	Расходы топлива на производство, ТУТ	9,802
01.07.2017	Расходы топлива на производство, ТУТ	8,575
01.08.2017	Расходы топлива на производство, ТУТ	10,308
01.09.2017	Расходы топлива на производство, ТУТ	15,041
01.10.2017	Расходы топлива на производство, ТУТ	57,012
01.11.2017	Расходы топлива на производство, ТУТ	76,084
01.12.2017	Расходы топлива на производство, ТУТ	79,691
01.01.2018	Расходы топлива на производство, ТУТ	92,665
01.02.2018	Расходы топлива на производство, ТУТ	103,249
01.03.2018	Расходы топлива на производство, ТУТ	105,474
01.04.2018	Расходы топлива на производство, ТУТ	59,521
01.05.2018	Расходы топлива на производство, ТУТ	13,938
01.06.2018	Расходы топлива на производство, ТУТ	8,466
01.07.2018	Расходы топлива на производство, ТУТ	9,189
01.08.2018	Расходы топлива на производство, ТУТ	9,775
01.09.2018	Расходы топлива на производство, ТУТ	14,412
01.10.2018	Расходы топлива на производство, ТУТ	55,871
01.11.2018	Расходы топлива на производство, ТУТ	80,362
01.12.2018	Расходы топлива на производство, ТУТ	98,427

Рисунок 3 - Выбор показателя

Далее открывается мастер обработки «Автокорреляция» и настраиваются параметры столбцов: поле «Дата» и «Показатель» неиспользуемые, а поле «Объем» используемое.

Предположим, что в данных имеет место сезонность и она не более года. В связи с этим зададим количество отсчетов равным 15 (тогда будет искаться зависимость от месяца, двух, пятнадцати месяцев назад). Диаграмма автокорреляционного анализа (АФК) представлена ниже на рисунке 4.

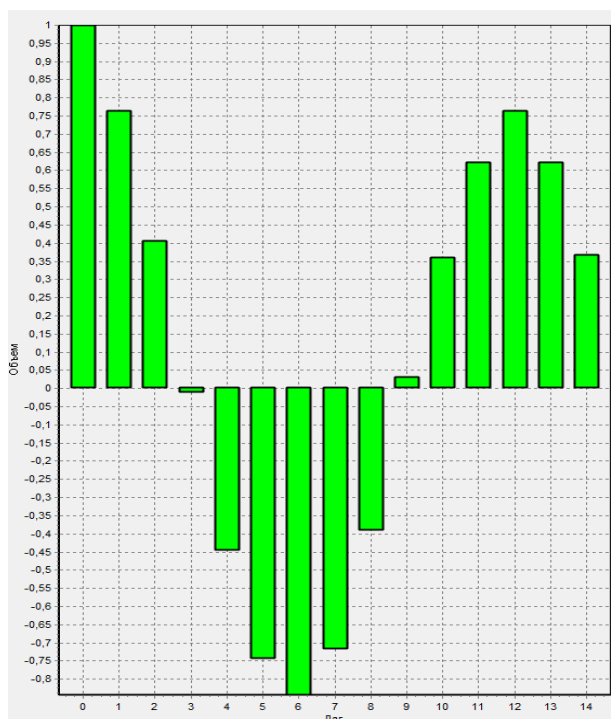


Рисунок 4 - Автокорреляционный анализ

Вначале корреляция равна единице – это означает, что значение зависит само от себя. Первый пик значений находится на 12 отсчете. Это соответствует годовой сезонности, оказавшемуся равным одному году.

Первое минимальное значение - на шестом отсчете и это значение близко к $-0,85$. Это так называемая обратная автокорреляционная зависимость. Данная зависимость не всегда присутствует в АФК. В нашем случае это зависимость соответствует половине периода сезонности.

Таким образом, в расходах ТЭПК присутствует явно выраженная тенденция сезонности. Если процесс носит сезонный характер, необходимо иметь данные как минимум за один полный сезон с возможностью варьирования интервалов.

На следующем этапе построения модели проводится декомпозиция временного ряда. С помощью мастера обработки «Декомпозиция временного ряда» составляется сезонная, трендовая и нерегулярная компоненты. Для этого в настройках мастера обработки для показателя «Объем» указывается тип сезонности - годовая.

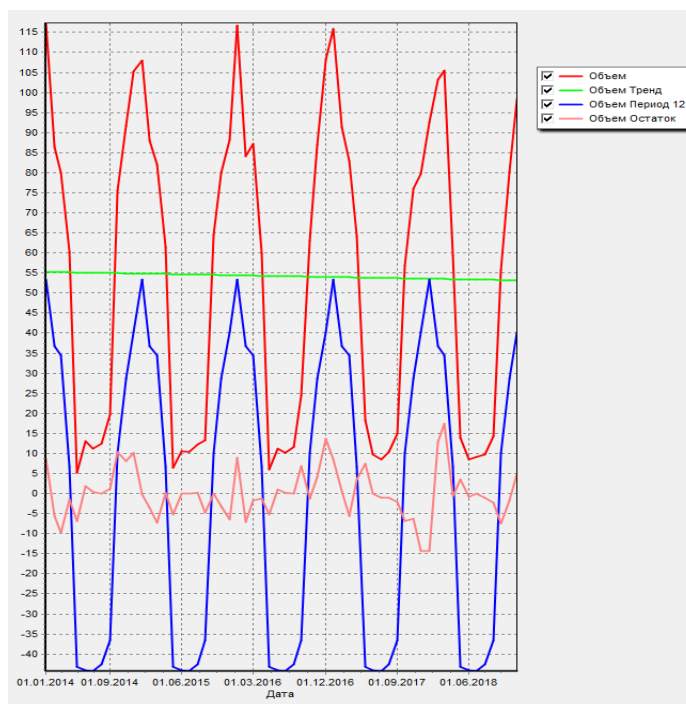


Рисунок 5 - Декомпозиция временного ряда

На четвертом этапе выполняется трансформация данных по методу скользящего окна. Строить прогноз на год вперед можно, основываясь на данных за 1 - 12 месяцев назад. Поле "Объем" выбирается как используемое, а глубина погружения равна 12. Результат работы мастера обработки представлен в виде таблицы на рисунке 6.

Дата	Показатель	Объем-6	Объем-5	Объем-4	Объем-3	Объем-2	Объем-1	Объем	Объем-12	Объем-11	Объем-10	0
01.01.2015	Расходы топлива на производство, ТУТ	11,236	12,552	19,634	75,529	91,728	106,351	108,098	117,368	86,395	80	
01.02.2015	Расходы топлива на производство, ТУТ	12,552	19,634	75,529	91,728	106,351	108,098	87,997	66,395	80	60,344	
01.03.2015	Расходы топлива на производство, ТУТ	19,634	75,529	91,728	106,351	108,098	87,997	82,148	80	60,344	5,216	
01.04.2015	Расходы топлива на производство, ТУТ	75,529	91,728	106,351	108,098	87,997	82,148	61,564	60,344	5,216	13,015	
01.05.2015	Расходы топлива на производство, ТУТ	91,728	106,351	108,098	87,997	82,148	61,564	6,387	5,216	13,015	11,236	
01.06.2015	Расходы топлива на производство, ТУТ	106,351	108,098	87,997	82,148	61,564	6,387	10,7	13,015	11,236	12,552	
01.07.2015	Расходы топлива на производство, ТУТ	108,098	87,997	82,148	61,564	6,387	10,7	10,413	11,236	12,552	19,634	
01.08.2015	Расходы топлива на производство, ТУТ	87,997	82,148	61,564	6,387	10,7	10,413	12,272	13,256	19,634	75,529	
01.09.2015	Расходы топлива на производство, ТУТ	82,148	61,564	6,387	10,7	10,413	12,272	13,256	19,634	75,529	91,728	
01.10.2015	Расходы топлива на производство, ТУТ	61,564	6,387	10,7	10,413	12,272	13,256	64,676	75,529	91,728	106,351	
01.11.2015	Расходы топлива на производство, ТУТ	6,387	10,7	10,413	12,272	13,256	64,676	80,08	91,728	106,351	108,098	
01.12.2015	Расходы топлива на производство, ТУТ	10,7	10,413	12,272	13,256	64,676	80,08	88,334	106,351	108,098	87,997	
01.01.2016	Расходы топлива на производство, ТУТ	10,413	12,272	13,256	64,676	80,08	88,334	116,702	108,098	87,997	82,148	
01.02.2016	Расходы топлива на производство, ТУТ	12,272	13,256	64,676	80,08	88,334	116,702	84,107	87,997	82,148	61,564	
01.03.2016	Расходы топлива на производство, ТУТ	13,256	64,676	80,08	88,334	116,702	84,107	87,207	82,148	61,564	6,387	
01.04.2016	Расходы топлива на производство, ТУТ	64,676	80,08	88,334	116,702	84,107	87,207	59,767	61,564	6,387	10,7	
01.05.2016	Расходы топлива на производство, ТУТ	80,08	88,334	116,702	84,107	87,207	59,767	5,949	6,387	10,7	10,413	
01.06.2016	Расходы топлива на производство, ТУТ	88,334	116,702	84,107	87,207	59,767	5,949	11,2	10,7	10,413	12,272	
01.07.2016	Расходы топлива на производство, ТУТ	116,702	84,107	87,207	59,767	5,949	11,2	10,174	10,413	12,272	13,256	
01.08.2016	Расходы топлива на производство, ТУТ	84,107	87,207	59,767	5,949	11,2	10,174	11,656	12,272	13,256	64,676	
01.09.2016	Расходы топлива на производство, ТУТ	87,207	59,767	5,949	11,2	10,174	11,656	24,551	13,256	64,676	80,08	
01.10.2016	Расходы топлива на производство, ТУТ	59,767	5,949	11,2	10,174	11,656	24,551	63,112	64,676	80,08	88,334	
01.11.2016	Расходы топлива на производство, ТУТ	5,949	11,2	10,174	11,656	24,551	63,112	86,953	80,08	88,334	116,702	
01.12.2016	Расходы топлива на производство, ТУТ	11,2	10,174	11,656	24,551	63,112	86,953	108,145	88,334	116,702	84,107	
01.01.2017	Расходы топлива на производство, ТУТ	10,174	11,656	24,551	63,112	86,953	108,145	115,842	116,702	84,107	87,207	
01.02.2017	Расходы топлива на производство, ТУТ	11,656	24,551	63,112	86,953	108,145	115,842	91,463	84,107	87,207	59,767	
01.03.2017	Расходы топлива на производство, ТУТ	24,551	63,112	86,953	108,145	115,842	91,463	82,925	87,207	59,767	5,949	
01.04.2017	Расходы топлива на производство, ТУТ	63,112	86,953	108,145	115,842	91,463	82,925	64,2	59,767	5,949	11,2	
01.05.2017	Расходы топлива на производство, ТУТ	86,953	108,145	115,842	91,463	82,925	64,2	18,347	5,949	11,2	10,174	
01.06.2017	Расходы топлива на производство, ТУТ	108,145	115,842	91,463	82,925	64,2	18,347	9,802	11,2	10,174	11,656	
01.07.2017	Расходы топлива на производство, ТУТ	115,842	91,463	82,925	64,2	18,347	9,802	8,575	10,174	11,656	24,551	
01.08.2017	Расходы топлива на производство, ТУТ	91,463	82,925	64,2	18,347	9,802	8,575	10,308	11,656	24,551	63,112	
01.09.2017	Расходы топлива на производство, ТУТ	82,925	64,2	18,347	9,802	8,575	10,308	15,041	24,551	63,112	86,953	
01.10.2017	Расходы топлива на производство, ТУТ	64,2	18,347	9,802	8,575	10,308	15,041	57,012	63,112	86,953	108,145	
01.11.2017	Расходы топлива на производство, ТУТ	18,347	9,802	8,575	10,308	15,041	57,012	76,084	86,953	108,145	115,842	
01.12.2017	Расходы топлива на производство, ТУТ	9,802	8,575	10,308	15,041	57,012	76,084	79,691	108,145	115,842	91,463	
01.01.2018	Расходы топлива на производство, ТУТ	8,575	10,308	15,041	57,012	76,084	79,691	92,665	115,842	91,463	82,925	
01.02.2018	Расходы топлива на производство, ТУТ	10,308	15,041	57,012	76,084	79,691	92,665	103,249	91,463	82,925	64,2	
01.03.2018	Расходы топлива на производство, ТУТ	15,041	57,012	76,084	79,691	92,665	103,249	105,474	82,925	64,2	18,347	
01.04.2018	Расходы топлива на производство, ТУТ	57,012	76,084	79,691	92,665	103,249	105,474	59,521	64,2	18,347	9,802	
01.05.2018	Расходы топлива на производство, ТУТ	76,084	79,691	92,665	103,249	105,474	59,521	13,938	18,347	9,802	8,575	
01.06.2018	Расходы топлива на производство, ТУТ	79,691	92,665	103,249	105,474	59,521	13,938	8,466	9,802	8,575	10,308	
01.07.2018	Расходы топлива на производство, ТУТ	92,665	103,249	105,474	59,521	13,938	8,466	9,189	8,575	10,308	15,041	
01.08.2018	Расходы топлива на производство, ТУТ	103,249	105,474	59,521	13,938	8,466	9,189	9,775	10,308	15,041	57,012	
01.09.2018	Расходы топлива на производство, ТУТ	105,474	59,521	13,938	8,466	9,189	9,775	14,412	15,041	57,012	76,084	
01.10.2018	Расходы топлива на производство, ТУТ	59,521	13,938	8,466	9,189	9,775	14,412	55,871	57,012	76,084	79,691	
01.11.2018	Расходы топлива на производство, ТУТ	13,938	8,466	9,189	9,775	14,412	55,871	80,362	76,084	79,691	92,665	
01.12.2018	Расходы топлива на производство, ТУТ	8,466	9,189	9,775	14,412	55,871	80,362	98,427	79,691	92,665	103,249	

Рисунок 6 - Трансформация данных к скользящему окну

Для построения прогноза с помощью метода линейной регрессии используется соответствующий мастер обработки «Линейная регрессия». Предположим, что на прогноз влияют данные за 1, 2, 4, 6, 7, 9 и 11 месяцев назад, тогда входными столбцами указываются эти поля. В качестве выходного поля выступает столбец "Объем".

На последнем шаге мастера обработки в качестве способа отображения выбираются диаграмма рассеяния. Как видно из диаграммы рассеяния, обучение прошло с высокой точностью, поскольку, ни одно из значений обучающей выборки не вышло за пределы границ. Данная модель подходит для осуществления прогноза ТЭПК.

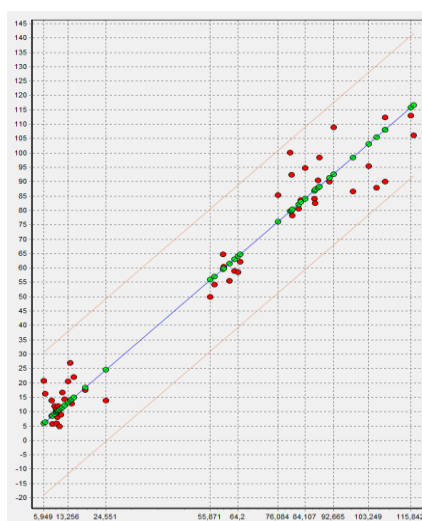


Рисунок 7 - Диаграммы рассеяния линейной регрессии

Для построения прогноза запускается мастер обработки – «Прогнозирование». На первом шаге мастера необходимо осуществить

настройку связи столбцов. Укажем связь между столбцами и горизонт прогноза, в нашем случае горизонт прогноза будет равен 12, так как построить прогноз необходимо на год вперед. Для наглядного представления полученных результатов используется диаграмма прогноза. Желтой линией отмечен построенный с помощью линейной регрессии прогноз.

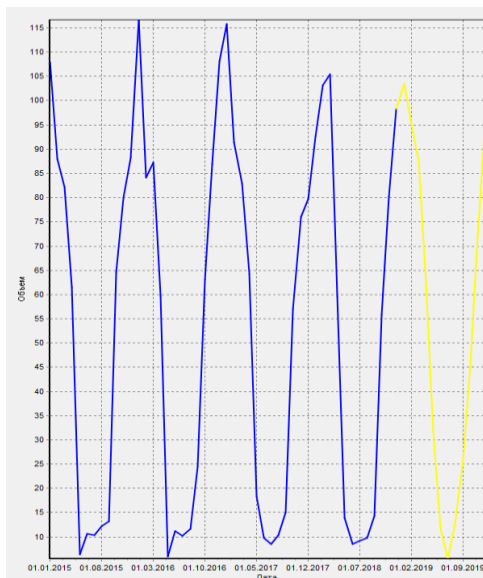


Рисунок 8– Диаграмма прогноза линейной регрессии

В завершение настраиваются отображаемые столбцы: Дата, Показатель, Объем и Шаг. На рисунке 9 представлен прогноз на будущий год.

01.09.2018	Расходы топлива на производство, ТУТ	14,412	
01.10.2018	Расходы топлива на производство, ТУТ	55,871	
01.11.2018	Расходы топлива на производство, ТУТ	80,362	
01.12.2018	Расходы топлива на производство, ТУТ	98,427	
01.01.2019	Расходы топлива на производство, ТУТ	103,474001546192	1
01.02.2019	Расходы топлива на производство, ТУТ	95,0314901027056	2
01.03.2019	Расходы топлива на производство, ТУТ	87,3906856889785	3
01.04.2019	Расходы топлива на производство, ТУТ	62,1965971718846	4
01.05.2019	Расходы топлива на производство, ТУТ	31,5015250736858	5
01.06.2019	Расходы топлива на производство, ТУТ	11,6221162389924	6
01.07.2019	Расходы топлива на производство, ТУТ	5,54219663125559	7
01.08.2019	Расходы топлива на производство, ТУТ	12,8112110388854	8
01.09.2019	Расходы топлива на производство, ТУТ	25,9506744921845	9
01.10.2019	Расходы топлива на производство, ТУТ	45,5636065846075	10
01.11.2019	Расходы топлива на производство, ТУТ	72,1146360556109	11
▶ 01.12.2019	Расходы топлива на производство, ТУТ	94,5340724129437	12

Рисунок 9 – Расчет прогноза с помощью линейной регрессии

Рассмотренный пример показал возможность использования линейного регрессионного анализа для прогнозирования ТЭПК. Об эффективности работы метода мы можем судить по таблице 2. По ней мы можно сделать оценку качества модели прогнозирования при помощи сравнения

фактических данных за январь, февраль, март 2019 года и результатов, полученных при помощи модели прогноза.

Таблица 2 – Сравнения фактических данных и результатов

Наименование показателя	Дата	Фактические расходы	Прогноз с помощью линейной регрессии
Ресурсы тепловой энергии	Январь 2019	607,637	599,04
	Февраль 2019	578,127	578,50
	Март 2019	521,166	524,07

Построенная с помощью линейной регрессии модель показывает хорошую точность. Работы по прогнозированию данных не заканчиваются при построении первой же модели, необходимо проводить регулярную верификацию результатов, для чего необходимо применять экспертные оценки, тестировать модель на данных, неизвестных при построении модели, строить их альтернативными способами.

Библиографический список

1. Интеллектуальный анализ данных: учеб. пособие для студентов специальности 080801.65 «Прикладная информатика (в экономике)» / Саратовский государственный социально-экономический университет. – Саратов, 2012. 92 с.
2. <https://studfiles.net/preview/2470562/>
3. Барсегян А.А. Куприянов В.В. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336с.