

Прогнозирование геопространственной доступности на основе больших данных компании AutoTel

Суворов С.В., профессор, кандидат экономических наук, преподаватель, заведующий кафедрой «Прикладная информатика» ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Царькова Н.И., кандидат педагогических наук, доцент кафедры «Прикладная информатика» ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Серякова А.О., магистрант, ФГБОУ ВО «Московский политехнический университет», Москва, Россия

Аннотация. В данной статье на практике показано применение машинного обучения к большим данным. В качестве Big Data используются данные компании AutoTel, включающие в себя более 6,8 миллионов наблюдений. Первоначальным этапом было проведение анализа больших данных, с целью получения базы для машинного обучения. После обработанные данные были сформированы в модель, и к ним применилось машинное обучение. Выходные данные были визуализированы на графике и карте.

Ключевые слова: машинное обучение, большие данные, Python 3, геопространственная доступность, анализ данных.

Prediction of geospatial availability based on Big Data of the company AutoTel

Suvorov S.V., Professor, Candidate of Economic Sciences, Lecturer, Head of the Department of Applied Informatics, Moscow Polytechnic University, Moscow, Russia

Tsarkova N.I., candidate of pedagogical sciences, associate professor of the Department of Applied Informatics, Moscow Polytechnic University, Moscow, Russia

Seryakova A.O., master of the Moscow Polytechnic University, Moscow, Russia

Annotation. This article shows in practice the application of Machine Learning to Big Data. AutoTel's database, which includes more than 6.8 million observations, is used as Big Data. The initial step was to perform Big Data analysis in order to obtain a basis for machine learning. After the processed data was formed into a model and machine learning was applied to them. The output was visualized on a graph and map.

Keywords: machine learning, big data, Python 3, geospatial availability, data analysis.

Введение

В целях сокращения количества автомобилей, находящихся в собственности, город Тель-Авив запустил проект совместного использования автомобилей под названием AutoTel. Пользователи сервиса могут зарезервировать автомобиль с помощью мобильного приложения и оплачивать его поминутно. Проект, который был запущен в октябре 2017 года, за один месяц привлек более 7500 пользователей, причем более 50% из них, теперь пользуются сервисом как минимум раз в неделю.

Для того чтобы услуга была надежной и удобной, AutoTel должен обеспечить, чтобы спрос и предложение были сбалансированы с географической точки зрения, то есть автомобили находятся там, где и когда они необходимы. Эта задача чрезвычайно сложна, так как автомобили управляются и припаркованы клиентами, которым не важна оптимизация, задуманная правительством. По большей части, распределение автомобилей не связано со спросом: одна из причин заключается в том, что, если автомобиль припаркован в пригородном районе, может пройти много времени, прежде чем

другой пользователь сможет доехать до центра города, где высокий спрос на машины существуют; таким образом, скопления неиспользованных автомобилей очень часто присутствуют на окраинах города [1].

Применив машинное обучение к большим данным компании, можно спрогнозировать геопространственную доступность автомобилей в определенный момент времени и использовать прогнозы для изменения бизнес-модели. Таким образом, AutoTel смогли бы, снизить цены на парковку автомобилей в зонах с повышенным спросом и планировать программу технического обслуживания, чтобы автомобили собирались в популярных районах.

Постановка цели и задач

Цель настоящей работы – анализ больших данных компании AutoTel с применением машинного обучения.

Для достижения поставленной цели были определены следующие задачи:

1. Очистка и подготовка данных для анализа.
2. Анализ основных показателей по территориальному и временному распределению доступных машин.
3. Подготовка данных для построения модели.
4. Применение машинного обучения к полученным данным.
5. Визуализация полученных данных.
6. Расчет основных экономических показателей компании.

Объектом работы является компания AutoTel.

Предметом исследования являются большие данные компании AutoTel по городу Тель-Авив, содержащие в себе 6,8 миллионов наблюдений. Данные включают в себя такие параметры, как: дата и время формирования заказа, точное место расположения автомобиля, общее количество машин, список машин.

Основной раздел

Построение и развертывание производственных моделей машинного обучения может быть довольно сложным. Даже с такими технологиями, как

Google Cloud Auto ML, Cloud ML Engine и другими готовые инструменты машинного обучения, обучающие модели и их использование в производственных системах обычно требуют обширного набора навыков, которые могут включать в себя различные передовые программы, Python, понимание сложных моделей, SQL и технологии баз данных. В данной статье продемонстрирована упрощенная технология обработки больших данных, основным инструментом которой выступает Python 3.

После проверки, очистки и сортировки данных, было произведено построение графика (рисунок 1), отображающего количество доступных машин по дням.

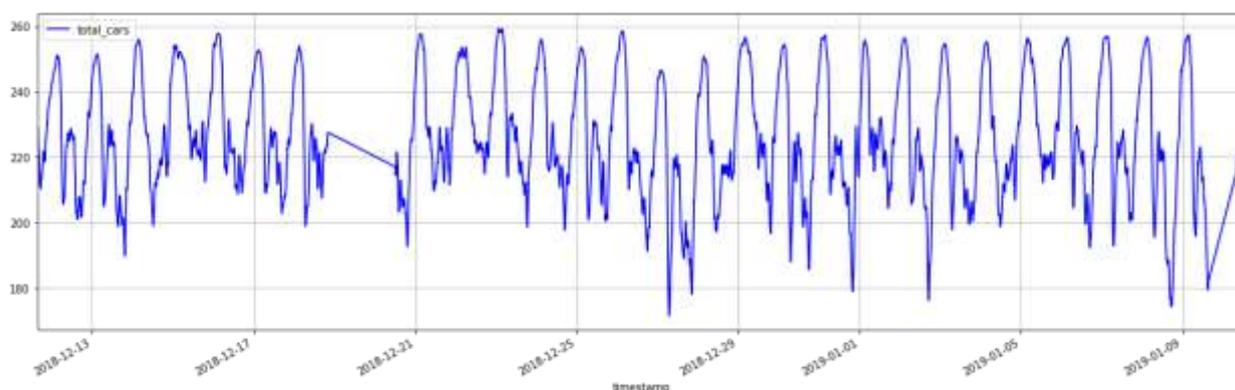


Рис. 1 – График доступных машин

Исходя из графика видно, что максимальное количество доступных машин равно 260, из чего можно сделать предположение, что это и есть примерное количество машин, зарегистрированных в сервисе. Следующим шагом было произведение расчета коэффициента использования машин. На полученном графике (рисунок 2), видно, что коэффициент использования машин варьируется от 2,5% до 25%, при этом пиком является 35%.

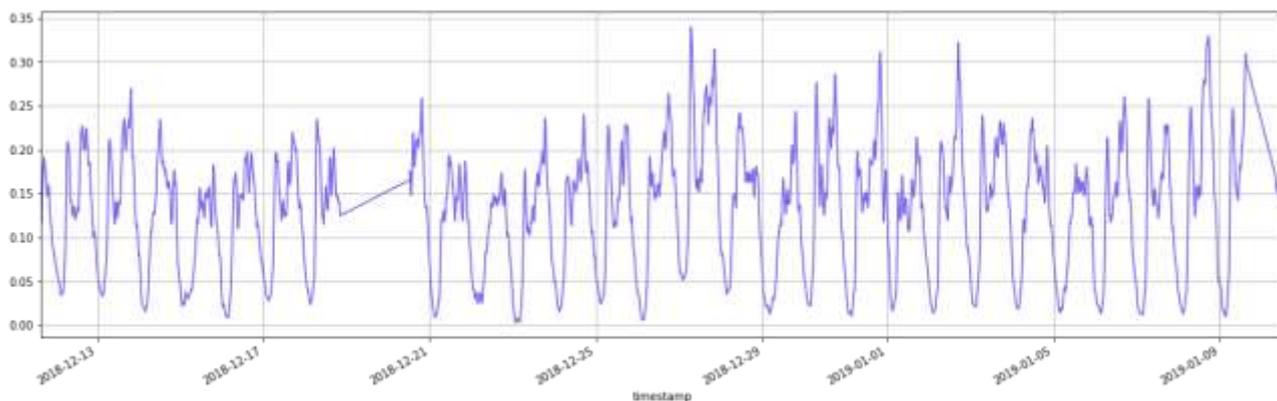


Рис. 2 – Коэффициент использования

При анализе тенденции использования (рисунок 3), можно заметить, что, в общем и целом, она растет.

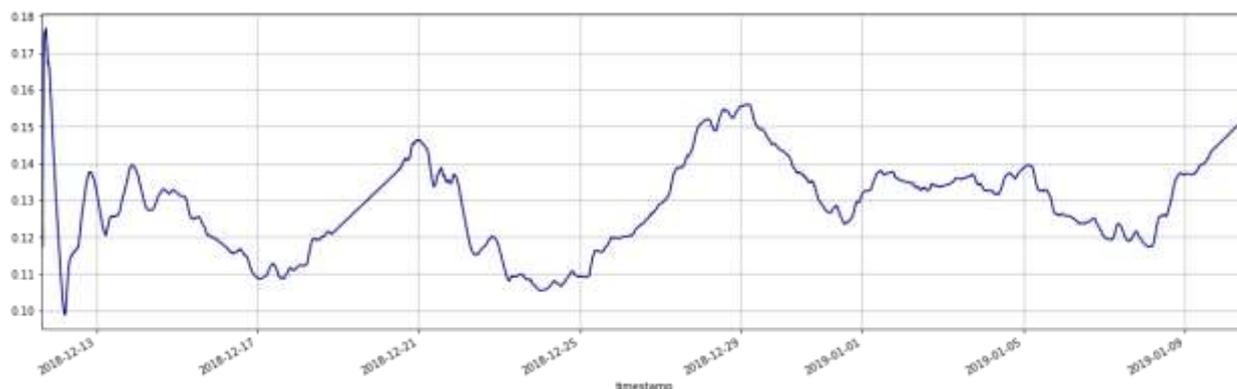


Рис. 3 – Тенденция использования

Далее был проведен анализ по часам и дням недели. Исходя из полученных данных можно сделать вывод, что пиком доступности машин является ночное время (рисунок 4). Именно в это время наибольшее количество свободных машин.

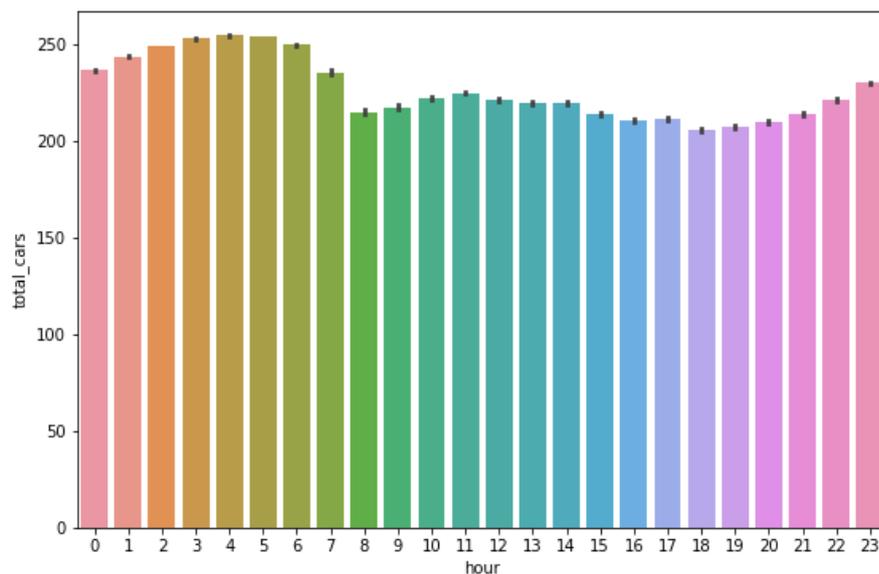


Рис. 4 – Доступность машин по часам

Выходные являются наиболее доступными днями (рисунок 5).

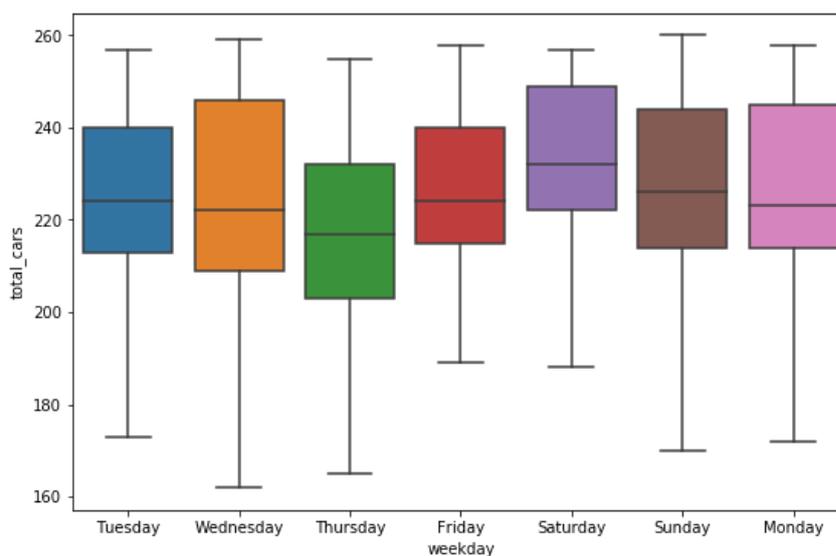


Рис. 5 – Доступность машин по дням недели

Анализ коэффициента использования подтверждает предположение, что наиболее популярными днями являются будни, вечернее время (рисунок 6). Пик приходится на 17-18 часов четверга. Это сочетание дня недели и времени суток является всемирно популярным как для услуг каршеринга, так и для сервисов заказа такси, таких, как например Uber [2]. Самыми популярными днями является четверг и среда (рисунок 7).

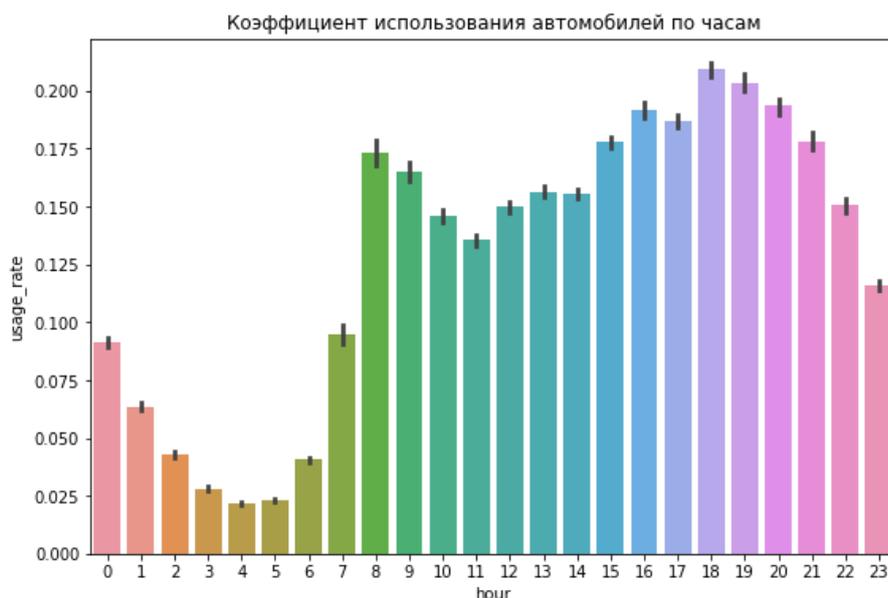


Рис. 6 – Коэффициент использования по часам

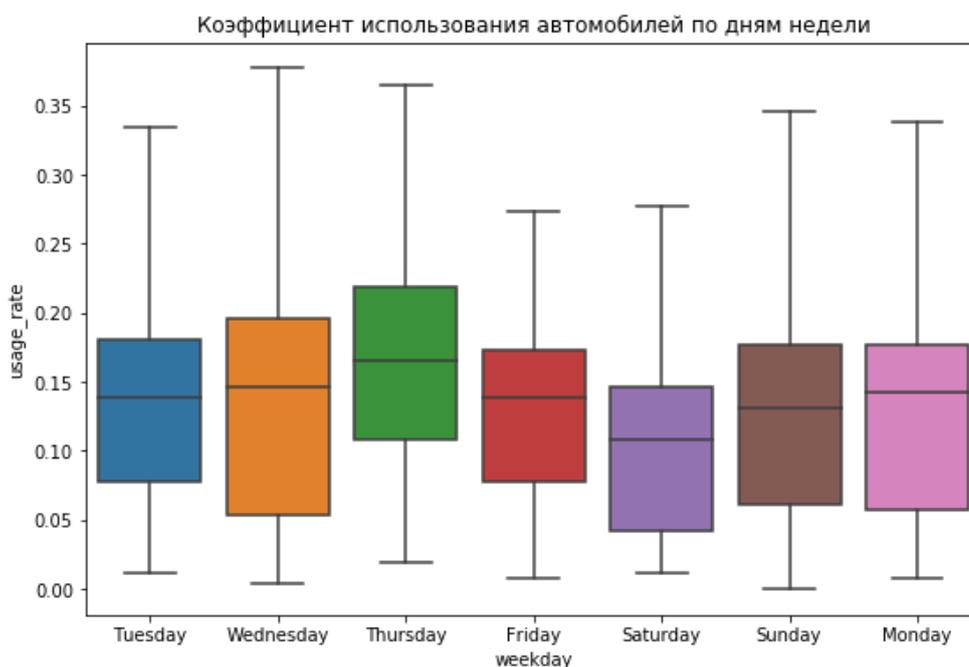


Рис. 7 – Коэффициент использования по дням недели

Трудность прогнозирования популярных точек использования автомобилей заключается в том, что нет четкого территориального деления на популярные и непопулярные районы, за исключением наличия повышенного спроса в центральных районах города (рисунок 8).

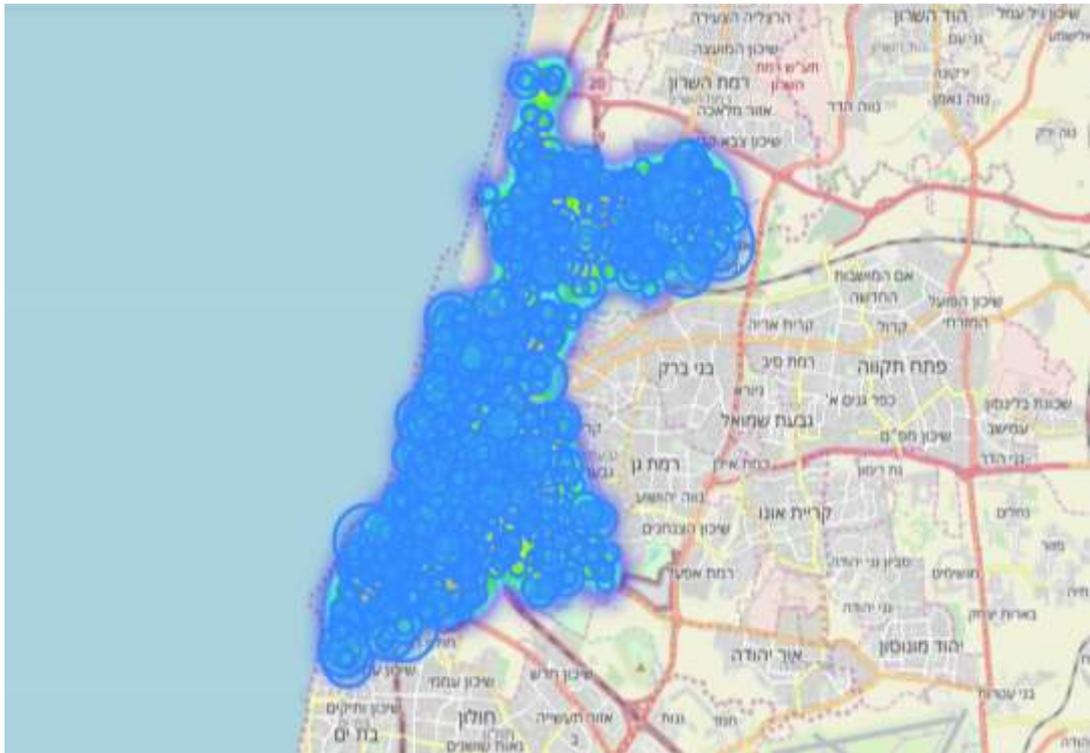


Рис. 8 – Территориальное распределение машин

Загрузив дополнительные данные по территориальному делению районов и дополнительную информацию об окрестности города Тель-Авив, можно перейти к прогнозированию количества доступных автомобилей в каждом из районов города. Предварительно необходимо сгруппировать данные по районам, исходя из показателей долготы и широты расположения машины. На графике (рисунок 9) видно, наличие особо популярных зон, таких как прибрежный район Спринг Клифф (Spring Cliff), Олд Джаффа (Old Jaffa) – район с художественными галереями, ресторанами, театрами, музеями и ночными клубами, одна из главных туристических достопримечательностей Тель-Авива [3].

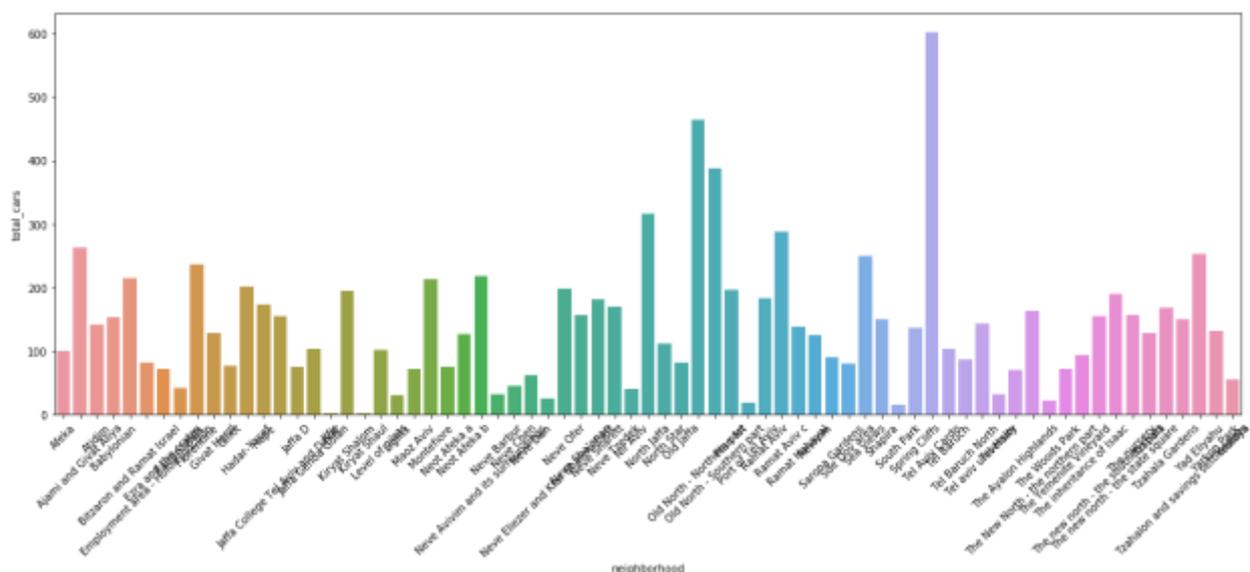


Рис. 9 – Распределение спроса на машины по районам

Следующим этапом было формирование модели для машинного обучения, путем категорирования основных признаков: коэффициента популярности района, дня недели и точного времени. Для обучения модели была определена выборка, включающая в себя наблюдения за одну неделю. По итогам обучения были получены приемлемые показатели по скорости обучения, среднеквадратичной ошибке, средней абсолютной ошибке и медиане абсолютной погрешности (рисунок 10).

```

evaluation:
r2_score: 0.5642690968881865
mean_absolute_error: 24.403850582740255
mean_squared_error: 1201.7824172726587
median_absolute_error: 16.804848357650414

```

Рис. 10 – Показатели скорости и ошибок машинного обучения

Полученные результаты по геопространственной доступности автомобилей в каждом из районов города были затем применены к тестовой выборке. Общий график доступности машин и его прогнозные значения (рисунок 11) показывает, что, в общем и целом, прогноз является достоверным и предсказательная модель является корректной.

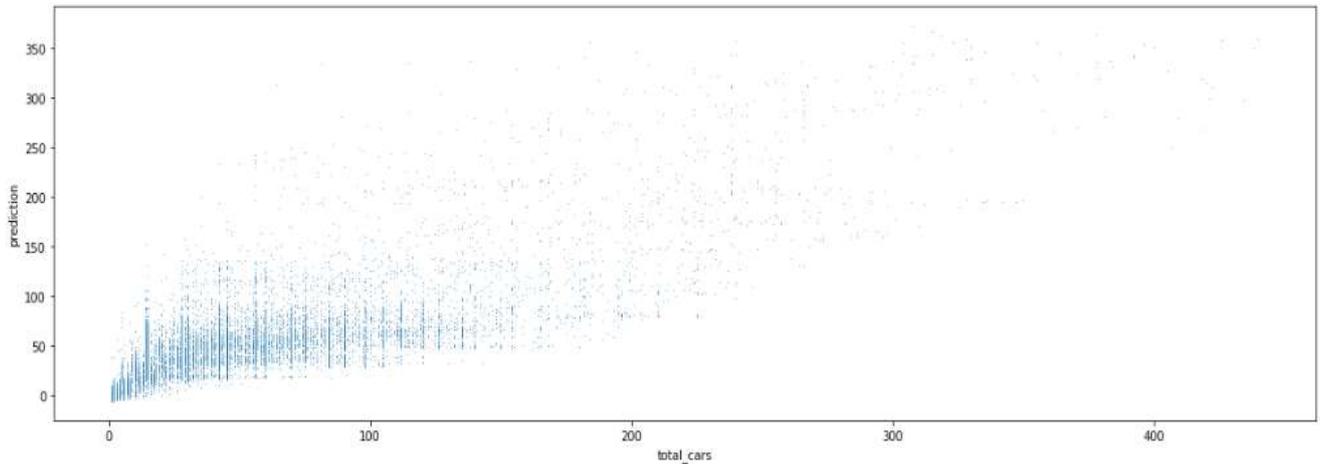


Рис. 11 – График прогнозных значений доступности машин

Для удобства восприятия данные были преобразованы в интерактивную карту, пример которой расположен на рисунке ниже (рисунок 12). Предполагаемая зона наличия спроса на машины в определенный интервал времени выделена цветовыми очагами.



Рис. 12 – Прогнозная карта доступности машин

Проанализировав данные длительности поездок (рисунок 13), был произведен расчет средней прибыли компании AutoTel за день, учитывая, что стоимость одной минуты фиксирована и составляет 1,3 шекеля.

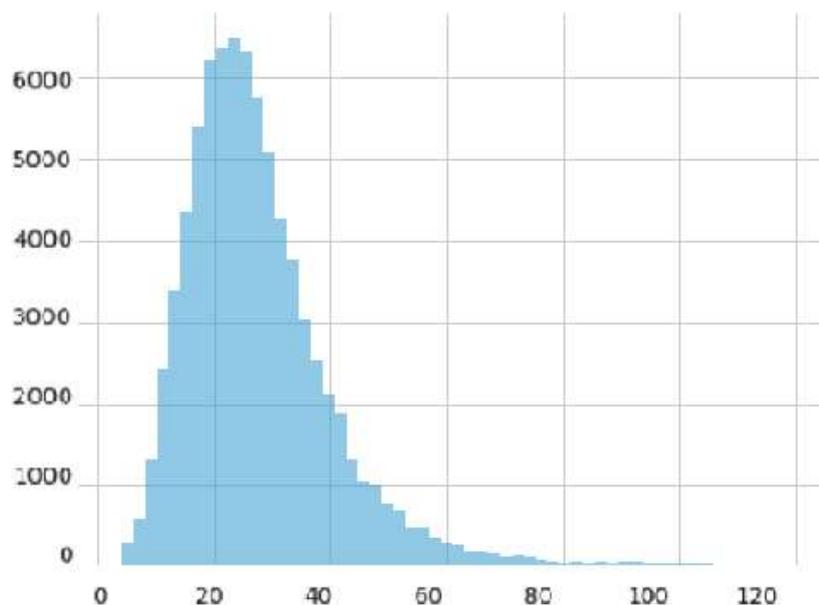


Рис. 13 – Средняя длительность поездок

Исходя из имеющихся данных можно построить общий график ежедневной прибыли компании (рисунок 14).

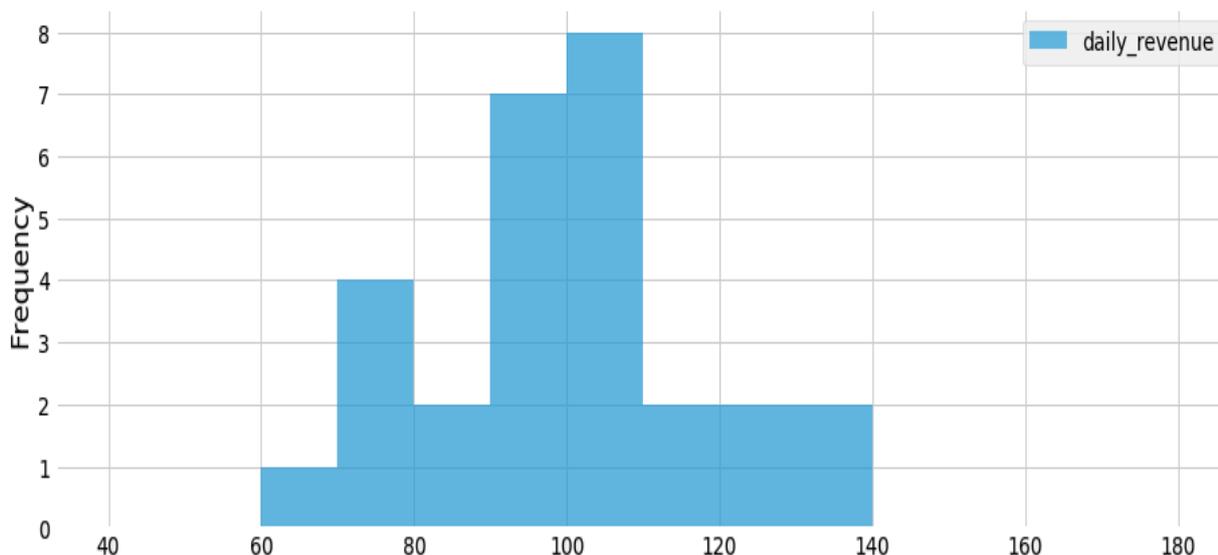


Рис. 14 – График ежедневной прибыли компании

Таким образом, исходя из полученных данных, средняя выручка компании составляет 100 000 шекелей. [4] Применяв полученные результаты анализа и машинного обучения, AutoTel сможет не только увеличить свою среднюю прибыль, но и повысить коэффициент использования автомобилей, а,

следовательно, и количество пользователей, которые будут предпочитать покупке собственного автомобиля использование услуг каршеринга.

Заключение

Выбор обозначенной тематики исследования обусловлен растущей актуальностью проблематики больших данных (Big Data) и методологий их обработки, в частности машинного обучения (Machine Learning). [5] Цель и задачи, поставленные в работе, выполнены. Таким образом, применение машинного обучения к большим данным компании AutoTel поможет достаточно точно спрогнозировать спрос и предложение автомобилей в определенный момент времени и территориальному расположению. Данное решение поможет воплотить в жизнь первоначальную идею и цель сокращения количества транспорта на улицах города. Помимо этого, все варианты развития события являются выгодными для компании, а риски являются минимальными. [6]

Библиографический список

1. AutoTel — официальный сайт компании AutoTel [Электронный ресурс]. Режим доступа: <https://www.autotel.co.il/en/>, свободный - (дата обращения: 13.09.2019).
2. UECS — электронный научный журнал управления экономическими системами [Электронный ресурс]. Режим доступа: http://uecs.ru/index.php?option=com_flexicontent&view=items&id=5595e, свободный - (дата обращения: 13.09.2019).
3. Википедия — свободная энциклопедия [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/Tel-Aviv>, свободный - (дата обращения: 13.09.2019).
4. Vesty — новостной портал для жителей Израиля [Электронный ресурс]. Режим доступа: <https://www.vesty.co.il/articles/0,7340,L-4908205,00.html>, свободный – (дата обращения: 13.09.2019).

5. Царькова Н.И., Ерисов В.Д., Пекова Е.А. Технология Big Data как инструмент управления в межкультурной коммуникации // Управление экономическими системами: электронный научный журнал, 2019. № 7.

6. Суворов С.В., Царькова Н.И., Спиридонова А.К. Анализ больших данных компании Uber Technologies Inc с помощью технологии Data Mining // Управление экономическими системами: электронный научный журнал, 2019. № 7.

References

1. AutoTel — the official website of AutoTel [Electronic resource]. Access mode: <https://www.autotel.co.il/en/>, free – (accessed date: 09/13/2019).

2. UECS — electronic scientific journal of economic systems management [Electronic resource]. Access mode: http://uecs.ru/index.php?option=com_flexicontent&view=items&id=5595e, free – (accessed: 09/13/2019).

3. Wikipedia — the free encyclopedia [Electronic resource]. Access mode: <https://ru.wikipedia.org/wiki/Tel-Aviv>, free – (accessed date: 09/13/2019).

4. Vesty— news portal for residents of Israel [Electronic resource]. Access mode: <https://www.vesty.co.il/articles/0,7340,L-4908205,00.html>, free – (accessed: 09/13/2019).

5. Tsarkova N.I., Erisov V.D., Pekova E.A. Big Data technology as a management tool in inter-cultural communication // Management of economic systems: electronic scientific journal. – 2019. – № 7.

6. Suvorov S.V., Tsarkova N.I., Spiridonova A.K. Big Data Analysis by Uber Technologies Inc Using Data Mining Technologies // Management of Economic Systems: Electronic Scientific Journal. – 2019. – № 7.