

УДК 004.85

ПРИМЕНЕНИЕ АЛГОРИТМОВ MACHINE LEARNING ДЛЯ ПОСТРОЕНИЯ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ДАННЫХ СТРОИТЕЛЬНОЙ КОМПАНИИ ООО «ПИ ГЕОРЕКОНСТРУКЦИЯ»

Е.В. Бородина, Университет ИТМО, Санкт - Петербург
А.З. Арсеньева, Университет ИТМО, Санкт - Петербург

В последние годы компании все чаще внедряют технологии обработки больших данных (Big Data). С помощью использования методов машинного обучения (Machine Learning) появилась возможность обрабатывать огромные массивы информации с невероятной для человека скоростью, управлять большими данными, анализировать данные, делать прогнозирование и визуализировать результаты.

Введение. Основными задачами данной работы являются сбор и анализ данных существующей компании ООО «ПИ Геореконструкция», формирование концепции применения алгоритмов машинного обучения для построения модели прогнозирования данных, с целью дальнейшей оптимизации процессов.

Компанией ООО «Пи Геореконструкция» были предоставлены данные о закупке карьерного строительного песка одного из поставщиков. Набор данных содержит в себе информацию о стоимости и доставке различного количества строительного песка на склад за последние 10 лет (количество материала, даты поставки, стоимость выезда машины за каждую доставку, стоимость за определенное количество песка, стоимость за доставку определенного количества песка). Также были предоставлены данные о стоимости выезда машины и стоимости доставки, которые имеют тарифную сетку с процентной зависимостью от объема.

В большинстве случаев каждая строительная компания реализует несколько проектов одновременно. В ходе производственного процесса осуществляется постоянный сбор, предварительная обработка, сортировка данных, и работа с большим объёмом информации силами сотрудников.

Основная часть. В данной работе был произведен анализ предоставленных данных, предложена концепция по применению алгоритмов Machine Learning для построения модели прогнозирования данных и её реализация. При работе с данными были использованы библиотеки scikit-learn, NumPy, SciPy, Pandas, Matplotlib, Keras.

Для прогнозирования требуемого объема песка ежемесячной закупки необходимо проанализировать данные на ежемесячном уровне и суммировать колонку с объемом закупки песка. Для этого необходимо было агрегировать данные с помощью группировки данных по месяцам. Таким образом наши данные будут представлены в две колонки, содержащие конкретную дату и объем, так мы сможем видеть объемы закупки, которые нам нужны.

Для дальнейшей работы было произведено преобразование данных. Для начала была выполнена проверка данных на стационарность – был построен общий график закупок, на котором была отображена тенденция к росту на протяжении всего периода. Исходя из этого появилась возможность посчитать разницу в закупке песка в сравнении с предыдущими периодами.

Последующим этапом была подготовка данных для последующего построения модели прогнозирования. Для этого была смоделирована разница и построен график, на котором отображаются отклонения в разный период времени. Для того, чтобы спрогнозировать закупку песка, были использованы данные закупок за предыдущие периоды. В качестве примера был взят период «обработки предыдущих значений» равный 12. Затем мы подбирали модель. В нашем случае была использована модель линейной регрессии (OLS), и рассчитан скорректированный R-квадрат.

Заключительным этапом было построение модели и масштабирование на все данные. Одним из важных моментов в данном этапе было разделение данных на обучающие и тестовые наборы. Обучающий и тестовый наборы нам нужны для оценки качества модели нашего

прогноза. Для набора тестов были взяты продажи за последние 6 месяцев. В качестве инструмента для масштабирования использовался один из классов масштабирования MinMaxScaler. Затем наша построенная модель обучалась по алгоритму.

Далее был сделан прогноз (предсказание), на основе ранее разделенных данных – тестовых и обучающих. Для того, чтобы сравнить результаты, необходимо было сделать обратное преобразование.

Выводы. В данной работе коэффициент прогноза равен 82%. На основе полученных данных мы строим графики для того, чтобы еще и визуально могли сравнить, насколько точно система сошлась с нашим прогнозом. Для получения более высокого коэффициента прогнозирования можно добавить больше данных и сопутствующих факторов и попробовать переучить модель.

Благодаря результатам данного прогноза мы можем предложить варианты оптимизации закупок и сформулировать ряд гипотез. Например, если, вместо нескольких раз в месяц закупать песок всего 1 раз, учитывая, что мы знаем о минимальном объеме песка каждый месяц, и его тенденцию к постоянному росту каждый год, то мы можем сэкономить на фиксированной стоимости доставки, которая учитывается при каждой закупке и уменьшить цену за 1 куб. метр, так как объем 1 закупки будет достаточно большой, что позволит по тарифной сетке закупить его с меньшей ценой.