

УДК 004.62

## **ПРОГНОЗИРОВАНИЕ ПРОДАЖ ЗАПАСНЫХ ЧАСТЕЙ ГОРНО-МЕТАЛЛУРГИЧЕСКОГО ОБОРУДОВАНИЯ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ИХ ЗАКУПКЕ**

**Никитин Е.Д.** (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО».)

**Научный руководитель – кандидат технических наук, доцент Кашевник А.М.**

(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО».)

### **Аннотация.**

В данном НИР была поставлена задача найти оптимальное решение для прогнозирования запасных частей горно-металлургического завода. Были проанализированы различные методы и подходы (такие как: метод Кростона, нулевой прогноз, наивный прогноз, прогноз скользящего среднего и другие) к прогнозированию продаж товарных позиций с непостоянным и единичным спросом. Был изучен рынок коммерческих предложений по данной тематике, сделаны выводы касательно того, какой способ лучше всего подходит к решению данной задачи, а также была построена модель прогноза. В результате был написан программный код для создания матриц корреляции, которые позволили отобрать наиболее релевантные позиции для дальнейшего обучения и построения модели прогнозирования.

### **Введение.**

Целью данной НИР было провести анализ способов, методов и инструментов для прогнозирования продаж запасных частей. На основе полученных данных было необходимо предложить и разработать собственный метод прогнозирования, ориентированный на горно-металлургическую промышленность. Объектом исследования были обезличенные данные о продажах различных запасных частей. Актуальность исследования заключается в новизне использования методов прогнозирования непостоянного спроса в России.

### **Анализ схожих работ.**

В основе исследования лежит анализ научных статей и публикаций на заданную тему. Интересно, что в большинстве исследований применяются такие методы и подходы, как «распределение Вейбулла», «распределение Пуассона», «нулевой прогноз и наивный прогноз», «прогноз скользящего среднего», «прогноз экспоненциального сглаживания», однако в ходе анализа был определен метод Кростона как основной для прогнозирования непостоянного спроса. Метод Кростона разработан специально для построения тех моделей, данные для которых непостоянны и единичны.

Помимо анализа научных статей были также проанализированы существующие коммерческие решения. Были выбраны 4 основных коммерческих решения, которые (как указано у них на сайтах) специализуются в области как постоянного, как и не постоянного спроса.

Воспользовавшись пробным периодом, было выявлено, что из всех программ «Novo Forecast Enterprise» наиболее точно подходит для анализа имеющихся данных, так как в своей основе использует всё тот же метод Кростона, а также имеет надстройку для Excel.

Благодаря проведенному исследованию стало понятно, что для более четкого прогноза необходимо, чтобы был учтен каждый временной промежуток. Это означает, что имеющиеся данные необходимо дорабатывать, так как при отсутствии продаж за день/неделю/месяц не ставится ноль, а просто проскакивает сам временной отрезок.

К сожалению, дороговизна Novo Forecast Enterprise и малый пробный период не дали возможность протестировать все функции. Также не было возможности построить матрицы корреляции для более глубокого сравнения.

### **Анализ и обработка данных.**

В рамках НИР использовались обезличенные данные в формате Excel характеризующие продажи запасных частей. Данные включают в себя: номер сделки (Deal ID), тип товара (Type), статус сделки (Stage), номер клиента (Company ID), источник (Source), даты открытия и закрытия (предполагаемого закрытия) сделки (Created и Assumed close date соответственно), товар (Element ID), количество проданного товара (Q-ty) и единица измерения (UOM).

Далее была проведена работа с предоставленными данными: очистка, сортировка, фильтрация и т.д. Для удобства первичного анализа была построена сводная таблица, в которой учитывались только те товары, сделка по которым уже состоялась или же была заключена. Также были взяты товары, единицы измерения которых могут быть сопоставимы и пригодны для сравнения. Были выбраны те товары по которым есть минимальная статистика продаж (которых было продано более десяти штук).

В среде разработки «Jupyter notebook» был написан программный код, который сначала открывает на чтение необходимый файл и нужный лист Excel и фильтрует данные по заданным ранее критериям.

Затем было необходимо извлечь из колонки «date» месяц и год, так как именно на основе данных о продажах за месяц будет считаться корреляция. Год нужен для того, чтобы условный январь 2019 и январь 2020 были разными числами и не суммировались в будущем. Для этого была написана отдельная функция «addYearWeekDay». Чтобы убедиться в правильности работы методы были выбраны наиболее коррелирующие товары и построен график по ним.

Следующим шагом было построение матрицы корреляции по покупателям, взяв за основу один товар. К сожалению, оказалось, что данных недостаточно, чтобы построить такую матрицу. Получилось, что максимальное количество продаж одного товара одному покупателю не больше трех.

### **Выводы.**

В ходе анализа полученных данных было определено, что большинство коммерческих решений неприспособлены для прогнозирования редких продаж. Также было отмечено, что имея более двух тысяч товарных позиций, для анализа подошли только 50, так как у них было десять и более продаж, а также более 30 проданных штук этих товаров. Тем не менее на основе этих результатов планируется построить модель для прогнозирования непостоянного спроса, которая будет дорабатываться в будущем.

Работая с данными о редких продажах, оказалось, что большинство имеющихся средств не подходят для анализа, так как информации о продажах совсем мало и в силу ошибок выгрузки и неполных данных о продажах, приходится фильтровать все данные, неподходящие для анализа. Фильтрация необходима для более наглядной, репрезентативной выборки, без каких-либо ошибок, такие как пустое поле покупателя, неправильная дата и прочее. Для того, чтобы можно было проверить правильность работы алгоритмов подсчета генерировались случайные значения, удовлетворяющие условию редких продаж. В дальнейшем планируется отработать уже имеющиеся алгоритмы на большем количестве данных, разработать модели и повышать её точность предсказания с помощью использования таких факторов как сезонность, нац. праздники и т.д.