

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА РЫНОЧНОЙ СТОИМОСТИ ДРЕВЕСИНЫ

Потапов Н.А.¹, Ярцева Н.А.², Ярцев М.Д.¹

Научный руководитель – канд. техн. наук, доцент Карманов А.Г.²

¹ СПб ГЛТУ им. Кирова

² Университет ИТМО

karmanov.nip@gmail.com

Введение

Разработана информационная система для анализа и прогнозирования рыночной стоимости древесины. Система автоматизирует сбор данных (веб-скрапинг), ETL-обработку, хранение, анализ и визуализацию информации о сделках, породах древесины и регионах. Цель – повышение обоснованности решений в лесопромышленном комплексе.

Основная часть

Информационная система реализована в виде микросервисного решения: серверная часть выполнена на Python/FastAPI, клиентская — на React/TypeScript, хранение данных организовано в PostgreSQL, а регулярные загрузки и обработка данных управляются через Apache Airflow. Такое разделение позволяет независимо развивать модули, масштабировать нагрузку и поддерживать устойчивую работу при росте числа источников и объема сделок.

Сбор данных о рынке древесины автоматизирован с помощью веб-скрапинга (Scrapy) и ориентирован на получение параметров сделок (цены, объемы, породы/сортименты, региональные признаки, временные метки). Далее в ETL-контуре выполняются очистка и нормализация: устранение дубликатов, приведение единиц измерения и форматов дат, унификация справочников пород и регионов, базовые проверки качества и согласованности записей. После загрузки в БД формируются подготовленные представления (витрины) для ускорения аналитических запросов и расчета показателей.

Аналитический блок включает вычисление агрегатов по породам, регионам и периодам (в т.ч. средние и средневзвешенные цены), анализ динамики и построение распределений. Для прогнозирования стоимости используются ML-модели RandomForest и XGBoost, обучаемые на исторических данных с учетом временных и категориальных признаков. В веб-интерфейсе доступен интерактивный дашборд с фильтрами, графиками изменения цен, гистограммами распределений и отдельным блоком прогнозирования, позволяющим получить ожидаемую стоимость по заданным параметрам и сопоставить ее с историческими значениями.

Выводы

Система прошла комплексное тестирование (функциональное, нагрузочное, безопасность) с успешностью 99%. Пилотное внедрение показало сокращение времени подготовки отчетов на 75%, точность прогнозной модели 94.2%. Расчет экономической

эффективности для масштабирования на 10 подразделений: ROI 339% за 3 года, срок окупаемости 2.3 месяца. Система готова к промышленному внедрению.

Литература

1. Лесной кодекс Российской Федерации: федеральный закон от 4 декабря 2006 г. № 200-ФЗ (ред. от 24 февраля 2025 г.). – Москва, 2006.
2. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning with Applications in Python. – New York: Springer, 2023. – 600 p.
3. VanderPlas J. Python Data Science Handbook: Essential Tools for Working with Data. – Sebastopol: O'Reilly Media, 2022. – 500 p.
4. Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45, № 1. – P. 5–32.
5. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794.
6. Гусаров В.И., Комаров А.С. Лесная таксация и лесоустройство. – М.: ГОУ ВПО МГУЛ, 2010. – 480 с.
7. Пилипенко А.В. Экономика лесного комплекса. – СПб.: Лань, 2018. – 320 с.
8. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. – Sebastopol: O'Reilly Media, 2022. – 550 p