DRC SYSTEMS

Driving Efficiency in Pharma Product Management with an AI-Powered Category Classification Solution

Client Overview

A leading global organization in the healthcare and pharmaceutical sector needed a scalable solution to classify thousands of unstructured product descriptions—spanning medical devices, pharmaceuticals, consumables, and supplements—into accurate, predefined categories. This classification was critical to support valuation processes, regulatory compliance, and streamlined operational workflows.

Business Challenge

The client faced a major bottleneck in managing and categorizing large volumes of product data stored in spreadsheets. The descriptions were highly varied, unstandardized, and included a mix of technical, commercial, and medical terminology. Manually processing such a vast dataset was not only time-consuming but also prone to errors and inconsistencies, significantly impacting data reliability and operational efficiency.

Key issues included:

Unstructured and Ambiguous Data :

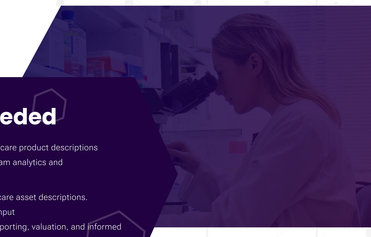
Product descriptions lacked consistent language and formatting, complicating manual classification.

High Dependency on Manual Sorting :

Human classification was slow, error-prone, and resource-intensive.

Need for Scalable Automation :

The client sought to automate classification across multiple asset types with high accuracy to support downstream analytics.



What Client Needed

To automatically classify unstructured healthcare product descriptions to reduce manual sorting, improve downstream analytics and compliance.

- Accurately categorize unstructured healthcare asset descriptions.
- Reduce manual effort and increase throughput
- Provide consistent tagging for improved reporting, valuation, and informed decision-making

What We Built

DRC Systems, a leading AI/ML development company, designed and delivered a **custom Machine Learning classification model,** integrated into a web application, to automate the categorization of healthcare and pharmaceutical data with high accuracy.

Solution Features:

Model Training & Evaluation

Applied and benchmarked Support Vector Machine (SVM) and Random Forest classifiers using labeled historical data.

Natural Language Processing (NLP)

Used NLP models to process, clean, tokenize, and vectorize medical and commercial terms for model training.

OCR Integration

Incorporated EasyOCR and Pytesseract for processing image-based product records and extending coverage beyond text-based data.

Web Interface for Review & Export

Delivered a simple, intuitive, and highly interactive web application enabling data upload, classification visualization, manual review, and export.

Business Impact

Classification Accuracy

Achieved **89% model accuracy** across pharmaceutical data

Operational Efficiency

Reduced manual sorting effort by over **70%**

Faster Data Processing

Enabled large-scale classification in minutes

Improved Decision-Making

Delivered clean, categorized data for reporting and analytics

Tech Stack

ML Algorithms for Classification

Support Vector Machine (SVM), Random Forest

Text Preprocessing

Python (Pandas, NLTK), Scikit-learn

OCR Integration for Image-Based Data

Pytesseract, EasyOCR

Web Application for Review and Export

Python with Flask

Conclusion

The solution enabled rapid scaling of product data management across departments and set the foundation for broader AI adoption. By automating the categorization of complex pharmaceutical product descriptions, DRC Systems significantly streamlined operational workflows, minimized manual effort, and enhanced the accuracy and accessibility of enterprise data across departments. The AI-powered accurate classification also helped the client support regulatory and internal compliance efforts.