DKomplex: DKODE- Intelligent Pricing Forecast Model

Team Members

Chika Starks, Daniel Arias, Deborah Shaw, Joshua Okero, Nicole Al-Sabah

Mentor

Justin Loza

Problem Statement

> *"How can DKomplex predict optimal pricing for plastic products without consistent data schemas and competitor information?"*

DKomplex is facing challenges with inconsistent data schemas across quotes, opportunities, and orders systems.

Data quality issues include missing values, inconsistent naming conventions, and internal sales that compromise historical pricing integrity.

The lack of usable competitor data can make it difficult to determine appropriate market pricing and optimize profitability.

Our Approach

Data Cleaning & Preparation

Clean historical data by enforcing consistent naming conventions and handling missing values. Convert records into consistent schemas for quotes, opportunities, and orders.

Feature Engineering

Integrate timing features like economic factors. Generate synthetic competitor data through data augmentation to enhance the model's predictive capabilities.

Model Development

Build regression models (SVR, tree-based) to predict optimal Price Per Unit (PPU). Evaluate using statistical metrics like R^2 score and cross-validation.

Visualization & Documentation

Use Matplotlib and Seaborn to interpret model performance. Create comprehensive documentation for both technical and non-technical audiences.

Agile Methodology

Weekly meetings, milestone reviews, and regular client feedback to ensure project stays aligned with business needs and adapts to changing requirements.

Dependencies & Risks

Dependencies

Technical Dependencies

- Secure access to DKomplex data systems
- Python environment (pandas, NumPy, PySpark)
- Fabric AutoML

Process Dependencies

- Data cleaning before model development
- Successful completion of foundational tasks
- Client feedback at milestone reviews

Risks

High Priority

- Delays in receiving/accessing DKomplex data
- Data quality issues (missing, incorrect data)
- Privacy/security concerns with confidential data

Medium Priority

- Limitations of synthetic competitor data
- Technical issues with environment
- Changes in project requirements/scope creep

Mitigation Strategies

Early communication with client, robust cleaning processes, NDAs for all team members, regular backups, and clear documentation of all assumptions.

Deliverables

Final Project Deliverables

Our deliverables provide DKomplex with immediate value and foundations for future expansion of their pricing strategy.

Cleaned & Integrated Dataset

Complete with code and documentation for future use and expansion.

Machine Learning Model(s)

Trained regression models for PPU prediction with evaluation summaries.

Visual Dashboard/Report

Key findings, model results, and feature importance analyses.

Technical Documentation

- Data pipeline documentation
- Code documentation with comments Methodology explanation
- Guide for future expansion

Final Presentation

- Presentation to DKomplex and advisors
- Summary of results and recommendations
- Client sign-off/approval

Future Team Handoff

- Lessons learned documentation
- Suggestions for future improvements

Current State

Milestones Achieved

Project Planning

Conducting literature review, setting project goals, and establishing environment access.

In Progress (Week 1-2)

Data Collection

Planning data extraction strategy and schema alignment approach for efficient processing.

Not Started (Week 2-3)

Feature Engineering

Planning for economic features and synthetic competitor data strategy to enhance model accuracy.

Not Started (Week 4-5)

Current Activities

- Reviewing literature on machine learning for price prediction
- Establishing communication channels with DKomplex stakeholders
- Securing necessary NDAs and access permissions
- Understanding previous team's contributions and challenges
- Analyzing existing pricing strategies and historical data

Initial Progress

- Completed project proposal with detailed goals and deliverables
- Identified key research papers on similar pricing models
- Established team roles and responsibilities
- Created initial project timeline and risk assessment

Project Timeline

Project Status: Just Beginning (Weeks 1-2)

Week 1-2	Week 2-3	Week 4-5	Week 5-7	Week 7-8	Week 9-10
Project Planning	Data Collection	Feature Engineering	Model Development	Evaluation	Final Delivery
In Progress	Not Started	Not Started	Not Started	Not Started	Not Started

Project Execution Approach

Our project follows an iterative approach, with clear milestones and deliverables at each phase.
We're currently in the initial planning stage, focusing on research and establishing project foundations. The timeline allows for appropriate time allocation to critical phases like data collection and model development.

Individual Contributions

Team Roles & Responsibilities

Chika

Feature Engineer

- Develops features based on economic factors
- Creates variable transformations
- Optimizes feature selection

Daniel

Evaluation Analyst

- Defines evaluation metrics
- Analyzes model output accuracy
- Prepares final reports

Nicole

Model Developer

- Reviews AutoML documentation
- Develops model training pipeline
- Packages model outputs and scripts

Joshua

Data Engineer

- Prepares raw data access
- Integrates external data sources
- Manages data pipeline documentation

Debora

Project Coordinator / Team Leader

- Coordinates team activities and inter-role synchronization
- Facilitates project kickoff and defines overall scope
- Manages project timeline and deliverables tracking