Project Name: DKomplex Intelligent Machine Learning Forecast Models

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

CST 499: Computer Science Capstone

Term: Summer 2025

Final Video Presentation Outline

- Introduction (~ 10 sec) - Deborah
    - In this project, we use machine learning models to predict sales forecasts for a client. We aim to identify the best-performing models and feature sets using Azure AutoML, followed by manual tuning to further enhance prediction accuracy. The models are evaluated using metrics such as R^2 scores, MAPE, and others to ensure reliability and performance. Our objective is to develop a robust forecasting solution that supports the client's strategic planning and decision-making.
- Problem Statement (~ 0.5-1 min) - Deborah
    - The client requires a reliable and accurate forecasting solution to predict future sales based on historical data and relevant external factors. Inaccurate forecasts can lead to overstocking, understocking, and poor strategic planning.
    - The goal of this project is to build and fine-tune machine learning models that can predict monthly sales quantities with high accuracy. Using tools like Azure AutoML and manual hyperparameter tuning, we aim to identify the most effective

models and feature combinations. The models will be evaluated using metrics such as R^2, MAPE, MAE, and RMSE to ensure robust performance and provide actionable insights for the client's decision-making process.

- Problem Solution (~15 sec) - Deborah
    - To address the client's need for accurate sales forecasting, we developed a machine learning pipeline that combines automated model selection with manual tuning. Using Azure AutoML, we identified top-performing models and feature sets. We then fine-tuned selected models to further improve performance
    - Technologies and tools Used:
        - Azure AutoML for model training, selection, and hyperparameter tuning
        - Azure Lakehouse for centralized data storage and processing
        - Python for data preprocessing and model evaluation
        - Pandas and Scikit-learn for data handling and metric computation
        - Evaluated forecast accuracy using R^2, MAPE, MAE, and RMSE
- Implementation Strategy (~ 25 sec) - Nicole

  Our implementation strategy focused on delivering accurate monthly quantity forecasts that DKomplex can use to calculate price per unit (PPU). We began by collecting and cleaning historical and external datasets, then engineered relevant features to improve prediction quality. Forecasting was performed exclusively through Microsoft Fabric's AutoML time series module, while other time series analyses and predictions were also explored manually to evaluate trends. These runs were conducted separately and compared for performance. Each group member was assigned weekly tasks and documented their progress in shared reports, which were reviewed during Tuesday and Sunday meetings with the DKomplex mentor. Based on feedback, we iteratively refined features, adjusted modeling parameters, and explored new data sources to improve forecast accuracy. The process emphasized reproducibility, collaboration, and strategic alignment with DKomplex's needs.

- Challenges (~ 25 sec) - Nicole

  Throughout the project, we faced several key challenges. First, many of the datasets provided were either incomplete or required significant cleaning and formatting before they could be used, especially for time series and forecasting. In some cases, the specific data DKomplex requested wasn't readily available, requiring us to source or engineer alternative features. Early on, learning how to navigate Microsoft Fabric and the AutoML UI was a learning curve for the group, the group was able to move forward with the AutoML tool after learning how to run notebooks in the UI. A major challenge was also understanding the structure and meaning of the data, which took time but was essential to making accurate predictions. Another technical limitation we encountered was that Microsoft Fabric only allowed one notebook to run at a time in AutoML, which created bottlenecks when multiple team members needed to test models on the same day. Coordinating around each other's schedules to share compute access added complexity to our workflow. Balancing model performance with interpretability and ensuring consistency across multiple sources made this a complex but valuable experience.

- Final Product Demo (~ 2 min) - Josh

  First, let's look at the data pipeline. All raw sales and external factor data is securely stored in Azure Lakehouse. For this example, we'll focus on the monthly sales data for one of DKomplex's key product lines.

  Next, we preprocess and transform the data using Python—handling missing values, normalizing features, and generating relevant variables. This cleaned dataset is then uploaded to Microsoft Fabric for model development.

Using Azure AutoML, we initiate a new time series forecasting experiment. The tool automatically evaluates dozens of algorithms across various feature combinations. Here, you can see the leaderboard: models are ranked by their R^2, MAPE, MAE, and RMSE scores.

After AutoML selects the best candidates, we extract those models for manual hyperparameter tuning. For instance, we adjusted lag features and retrained the model, which nudged our performance metrics higher.

To generate a forecast we input the most recent data and receive predicted monthly sales quantities for upcoming time periods. The system visualizes these predictions alongside actual historical figures, helping the DKomplex team spot trends, anticipate demand spikes, or prepare for potential downturns.

Finally, results and metrics are exported to Excel and Power BI so the client can interactively explore the forecasts. They can adjust key assumptions or filter by region or product, seeing updated predictions in real time.

Throughout this process, our team ensured that all steps are reproducible and well-documented—helping future teams, quickly pick up where we left off.

- Conclusion/Summary (~ 20 sec) - Chika

  In conclusion, we contributed significantly by adding 35 new datasets, 16 of which were approved, and engineering impactful features to improve forecasting accuracy. We tested these features across one-, three-, and twelve-month horizons, refining multiple models and finding strong performance from Extra Trees. Working closely with the DKomplex team, we strengthened the base model and delivered a robust forecasting framework, well-positioned for the next phase of development.

- Accomplishments: (~ 20 sec) - Chika

  Our team gathered new data, performed feature engineering, and tested various forecasting models to enhance the DKomplex base model. We explored different prediction horizons, used AutoML to evaluate algorithms, and fine-tuned configurations with DKomplex guidance, achieving improved forecasting performance and detailed results.

- Future Work: (~ 20 sec) - Daniel

  We've provided Excel reports and Microsoft Fabric notebooks to DKomplex with detailed insights and performance metrics that will enable the next group of students to build upon our findings without repeating any previous work. This allows them to focus more time on implementing the most promising model configurations and exploring more advanced techniques. DKomplex can use our research to continue developing their optimal PPU forecasting solution, with clear guidance towards future workers on which methods showed the most potential for further development.

- Reflection - Daniel

  This project provided valuable hands-on experience with enterprise-level machine learning tools and real-world forecasting challenges. Working with Microsoft Fabric and Azure AutoML taught us the importance of understanding both the technical capabilities and the limitations of automated ML platforms. The collaborative aspect with DKomplex's engineering team highlighted the significance of clear communication and iterative feedback in real-world projects. The experience of managing shared resources and coordinating team workflows in a cloud environment was also valuable for future industry work. Overall, this experience has demonstrated how successful data science projects require both technical proficiency and the ability to adapt to real-world challenges while collaborating effectively as a team.

- Closing slide: Thank you's (~ 10-15 sec) - Josh

  We'd like to thank the DKomplex engineering and management teams for their ongoing guidance and support throughout this project. We're also grateful to our CST 499 instructors and mentors for their insight and feedback. And of course, thank you to everyone watching our presentation today.

We appreciate your time and interest in our work. If you have any questions or would like to discuss any aspect of the project in more detail, please feel free to reach out.

Thank you!

5:30