Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

ROOT CAUSE ANALYSIS IN SUPPLY CHAIN PLANNING USING EXPLAINABLE MACHINE LEARNING

Pavle Kecman
Josephine Fang
Ana Glaser

NXP Semiconductors
6501 W. William Cannon
Austin, TX 78735 USA

ABSTRACT

In the highly dynamic world of semiconductor manufacturing, planning analysts are asked to analyze variations between weekly production plans with the goal of identifying a resolution in a landscape involving elaborate optimization models with significant interdependence between data elements. We propose a solution to effectively analyze the weekly planning engine output and identify the data elements with significant contribution to the outcome. An explainable Machine Learning model is trained and deployed to simulate the behavior of the planning engine. Each model execution can be explained to identify the features with the most significant contribution to prediction. The resulting application contributes to a timely resolution to the production plan deviation, while generating significant productivity gains.

1 INTRODUCTION

Semiconductor manufacturing is one of the most complex industrial processes. Manufacturers are supported by detailed supply chains and logistics operations plans to help meet the constantly increasing demand in a profitable manner.

Operations are planned using mathematical optimization and operations research techniques. A set of hierarchically executed LP problems are defined with the objective to minimize completion time of all received and forecasted sales orders. The LP models produce optimal manufacturing schedules by considering, among other factors, resource capacity, inventory levels of raw materials and semi-finished goods, and prioritization of sales orders.

The produced plans are sensitive to changes in internal and external factors. Internal factors are related to availability of resources and material and external factors include changes in demand and geopolitical dynamics. Given the high variability of these factors, the created plans can become obsolete quickly when deployed in a dynamic environment. The planning engine therefore operates as a time-driven open-loop controller. A production plan is executed on a weekly basis while planning (control) horizon is several years. High ratio between control horizon and sampling frequency may lead to instability of successive production plans.

2 PROBLEM DESCRIPTION

The variability of production plans represents a great challenge for supply chain and operations planners. Planners strive to identify the root cause of production plan changes and escalate to minimize the deviation from the committed deliveries.

Often, processes for root cause analysis are manual. A team of analysts manually inspect the tables in the enterprise data warehouse or available reports to determine the factor that, based on their expert opinion, is responsible for the changes in production plans. Potential productivity gains from automating this process amounts to 1 day of planner productivity per week, or 4 days saved per employee each month,

Besides being labor intensive, this manual process is also difficult, as analysts are constrained to assessing the factors individually. In reality however, the impact of a change of a single factor also depends on the values of other factors even if they remain constant. For example, the impact of material inventory level reduction for a particular part on production output is lower in situations of lower demand or lower capacity utilization. Moreover, the impact of the weekly fluctuations of LP engine inputs is often unevenly distributed across different product families. For these reasons Machine learning (ML) was identified as a potential solution to this root cause analysis (RCA) problem.

3 METHODOLOGY

Typical applications of ML for root cause analysis reduce the problem to a classification problem (Solé et al. 2017). Historical data, including the actual root causes, is used to train the model to be able to identify the root cause, given the values of predictive features that represent the current state. This methodology is not always applicable as there can be limited or no historical root cause data available.

The methodology we propose consists of the following steps:

1. Train a regression ML model to emulate the planning engine's behavior. The planning horizon is divided into quarters and the complex output of the planning engine is aggregated as the number of completed products per quarter. The target of the model is week-on-week delta in the number of completed products per quarter. The features of the model are week-on-week deltas of selected inputs to the planning engine such as: material inventory levels, resource capacity, order book and forecasted demand
2. Trained ML model can be used to predict the week-on-week deltas of feasible sales per product family
3. Each prediction of the ML model can be explained using Shapely values that quantify the contribution of each input feature to the model prediction. The method includes untangling the causality dependence graph of the potential root causes (Lundberg and Lee 2017).

4 COMPUTATIONAL RESULTS

The regression model was trained on 18 months of historical data comprising weekly plans for about 1500 product families. The model predicts feasible sales for each product family for the two upcoming quarters. Random Forest regressor (Breiman 2001) was selected as the ML model after applying a model selection process and hyperparameter optimization using grid search with cross-validation.

The trained model is deployed and subjected to weekly evaluations and user acceptance testing (UAT) after each planning engine run. R2 of the model was above 0.95 in 3 out of 4 weeks observed during UAT.

Planning analyst evaluations concluded that in ~75% of cases the root causes identified by the model (features with the highest Shapely values) match their expert opinion.

5 REFERENCES

Breiman, L., 2001. “Random Forests”. *Machine Learning*, 45, p. 5-32.

Lundberg, S.M. and S. I. Lee. 2017. “A Unified Approach to Interpreting Model Predictions”. *Advances in Neural Information Processing Systems*, 30.

Solé, M., V. Muntés-Mulero, A. I. Rana, and G. Estrada. 2017. “Survey on Models and Techniques for Root-Cause Analysis”. *arXiv* preprint arXiv:1701.08546.