End to end New Product Recommendation System Design Design for Six Sigma: Roadmap for Successful Corporate Goals, Salman Taghizadegan, in Essentials of Lean Six Sigma, 2006

Concept development phase	Technology Concepts Clients (VPs of Sales and Operations) expect an innovative Machine Learning- based New Product Recommendation System	Goals Maximize the return over the investment and revenue growth from new products	 3 years of past c 1 Data Scientist (at be released to production in 3 weeks ustomer journey history (part-time) and 1 ML Engineer (full time) uting environment with GPUs/TPUs
Design development phase	 Inputs Rolling window of historic customer journey transactional data (9 months for training and 3 months for validation) Social media feeds from Facebook, Instagram and LinkedIn 	Outputs Recommendation of products to be launched (0: not recommended 1: recommended	Flow Diagram(*) Data Privacy and Bia Customer and PII widata pipeline	a <u>s</u> Il be removed before used by model
Optimization development phase	Failure mode effects analysis (FMEA) • AUC - ROC curve to measure recommendation (classification performance) • Prediction accuracy/recall/f1-score	Random baseline: run random classifications on historical data Human baseline: gather performance of existing R&D and Marketing teams on new products recommendations Simple heuristic: gather voice of the customer and customer surveys		
 Verify capability and functionality phase	Tolerance design analysis			
Feasibility	Technical Feasibility Proof of Concept with top 100 products by sales volume will be conducted	Economic and Operation Feasibility confirmed by project budget and available company resources and specialized personnel		
Development	Data preprocessing & representation Historical data will be pre-processed to extract features of interest Features will be analyzed using Random Forest and OLS	Model Selection 1. Iterative process to select best model: 2. TPOT(Auto ML) for classifier based on quantit 3. Co-training for classifier based on social media		Training & Hyperparameter Tuning Feature Engineering A/B Testing
Preproduction	 Scaling Use data parallelism for training model pipeline (quantitative data) and co-training (social media feeds) 	 Serving Run experiments with Sales and Operations playend user experience, predictor behavior, bound results interpretability and model explainability Conduct ablation studies removing each communication 	laries, errors,	
Production		 keeping the rest to determine the efficiency of Find components whose removals eventually described reduce the model's performance but significant complexity. Check model for propagating any gender and refine if gender and race are features used) 	on't significantly ly reduce its	1

Customer Journey and Social Media Feeds Datawarehouse



Project Environment Feature Engineering (Random Forest & OLS) Numerical Quantitative Data Features

- Max Price
- Min Price
- Product category
- Marketing campaigns
- Promotions/discounts Sales volumes by product/category
- Voice of the Customer surveys
- Seasonality of products/categories

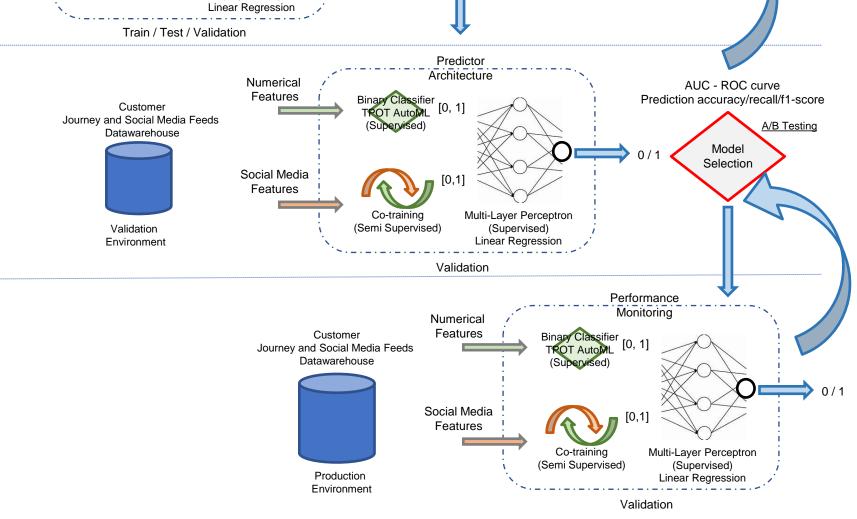
Social Media Feeds

- Positive/negative references to company brands/categories/products
- · Meta links to products websites

Social Media

Features

New Product Recommendation System Life Cycle



AUC - ROC curve

Prediction accuracy/recall/f1-score

Model

Selection

Predictor Architecture

[0,1]

Multi-Layer Perceptron

(Supervised)

Binary Classifier [0, 1]

Co-training

(Semi Supervised)