

CAPSTONE PROPOSAL Part 1

Project Overview	
Project Name	E-Commerce Returns and Profitability
Business Understanding*	<p><i>What problem are you trying to solve? What question are you trying to answer?</i></p> <p>What are the factors that influence buying behavior of e-commerce customers to minimize returns of products for businesses? We will first understand the feature importance of the factors within the given dataset, and we can give recommendations based on which factors influence customer returns. Those influential factors can be focused on by businesses so they can take action on those factors to minimize returns of customers, thus maximizing profits.</p>
Client/Project Motivation *	<p><i>Who is your client? What is the motivation behind your project?</i></p> <p>To help the e-commerce market industry gain insights on buying behavior to determine an effective way of increasing revenue. Focus on determining factors that lead to returns from customers.</p>
Market /Industry	<p><i>What market does your solution serve?</i></p> <p>E-Commerce businesses</p>
State of the Art*	<p><i>How has this problem been approached in the past?</i></p> <p>Many researchers are more focused on maximizing sales to boost profits from customer feedback and satisfactions rather than looking at specific factors that can be minimized in order to boost sales in the same way. In their cases, researchers stick to the positives and maximize those factors that keep the customer satisfied. In our case, prioritizing the idea of whether a customer returns a product can give information about an unsatisfied customer in which businesses can avoid those practices that makes the customer unsatisfied.</p>
Success Metrics (Evaluation)*	<p><i>What metrics will you use to determine success of your business recommendation?</i></p> <p><i>*cross-validation</i></p> <p>Prediction accuracy on returns, confusion matrix, and AUC-ROC to help us understand how well our model distinguishes between classes.</p>
Scalability*	<p><i>If successful, would it be possible to scale the project to size required to meet your client’s needs? What challenges do you anticipate?</i></p> <p>There is a possibility of being able to scale the project to size required to meet our client’s needs. The challenges could be the continuation of an unbalanced data set, maybe it could be even more unbalanced as the data set gets bigger. Since the current model will be trained by a dataset that contains unbalanced data, another challenge may be to retrain the model as the scale of the project continues to grow.</p>

CAPSTONE PROPOSAL Part 2

Modeling Requirements

Data Type (e.g., numerical, categorical)	<p><i>What type of data is needed to answer the posed question(s)?</i></p> <p>Target: categorical</p>
Data Source	<p>websites_e-commerce data set on Kaggle https://www.kaggle.com/datasets/willianoliveiragibin/websites-e-commerce Features: accessed_From, age, gender, country, membership, pay_method Target: Returns</p>
Data Preparation Steps*	<p><i>Which tools or methods are needed to prepare the data before using it?</i></p> <p>Under/over sampling, remove null values</p>
Data Challenges*	<p><i>Are there any issues, for example missing data, noise, class imbalance?</i></p> <p>Some columns (Age, Gender) have null/meaningless values, class imbalance</p>
Modeling Techniques*	<p><i>Indicate the selected algorithms and the reasoning behind the selection.</i></p> <p>Logistic Regression for a simple and baseline model representation, we will do more research to see if this linear decision boundary is best for our dataset or if there is another that could be more fit for our business problem. Random Forest</p>
Target Variable*	<p><i>Indicate the target variable and connect the choice to the proposed problem.</i></p> <p>Customer returns By identifying which demographics of customers return products more often we can advise the client on where to spend less resources on marketing.</p>
Regression or Classification problem	<p>Classification</p>
Tools/Methodologies	<p><i>Indicate the tools you plan to use to solve the problem. Name specific Python libraries or particular techniques that will be used.</i></p> <p>Sklearn, seaborn, matplotlib, pandas</p>

Project Timeline

Key:

- Complete
- On Track
- At Risk
- Off Track

					1/12	1/19	1/26	2/2	2/9	2/16	2/23	3/1	3/8
Week Number	Status Update	1	2	3	4	5	6	7	8	9	10	11	12
Milestones				Mentor Status Update	Proposal Due	VT Feedback Available	Flatiron Feedback Available	Mentor Status Update		Mentor Status Update		Practice Pres.	Final Pres.
Formulate business problem													
Find data set													
Data prep													
Modeling													
Evaluation/ Testing													
Final Results													
Final Results Review													
Presentation Prep													