COSC 202: MODELING AND SIMULATION

BSCS 2-*4* A.Y. 2023-2024

GROUP 6

MEMBERS:

- *CAEDO, Juliana*
- *FRILLES, Roljohn C.*
- *MAHAYAG, David*
- *MANCILLA, Nathalia*
- *OSANA, Lester Fuentes*
- *PLACENTE, Yesa*

PROJECT TITLE:

Natural Language Processing Model for Distinguishing Bot-Generated and Human-Created Shopee User Feedback in English and Filipino.

PROBLEM:

The accuracy of classifying bot-generated and human-created reviews on Shopee within English and Tagalog language poses significant complexity. This problem involves the increasing prevalence of automated bots generating reviews that could potentially damage the authenticity and reliability of user feedback from real customers' honest opinions and experiences with the product or service.

Furthermore, the rise of automated bots generating fake feedback affects the credibility of user feedback in the modern age of online reviews. The problem revolves around the question of how to solve the problem of identifying between human-created reviews and computer-generated information. The researchers aim to achieve this by creating a more trustworthy and honest online environment, preserving credibility to the authenticity of user experiences shared on digital platforms.
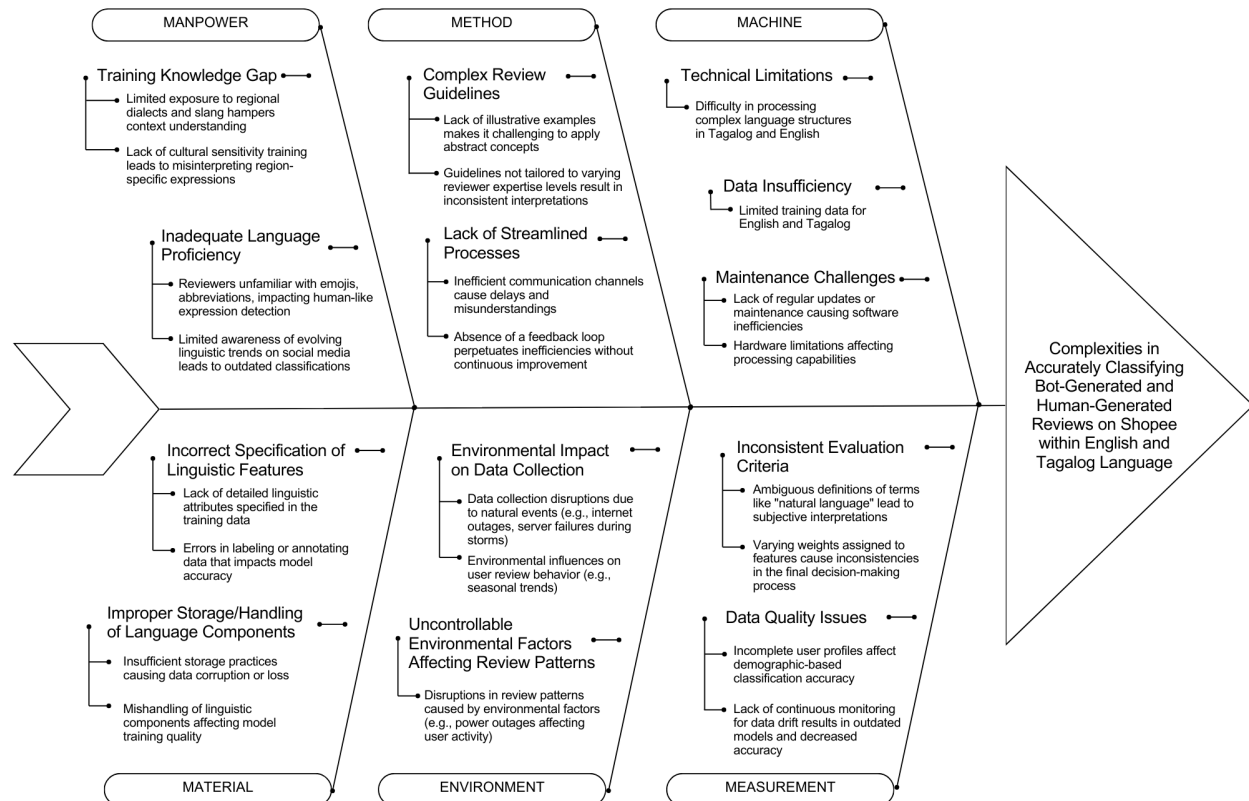
PROBLEM STATEMENT:

Shopee is the leading eCommerce and shopping application in the Philippines according to SimilarWeb's November 2023 data. Its popularity stems from its user-friendly interface, a wide range of trusted merchants, and huge monthly discounts, making it the preferred platform for diverse consumers seeking reliability and a hassle-free shopping experience. However, the platform is not capable of detecting any bot-generated reviews under its "Product Ratings" feature. Consequently, many users fall victim to false advertising and scams within the application. In such cases, the product they receive after delivery may work differently than described in the product description, or even worse, they may receive a completely different product altogether.

A recent report found more than one-third of Filipinos surveyed had been scammed or encountered retail fraud online (Royandoyan, 2023). In an advisory, the CICC said that the shopping scam rate in the country reached 35.9 percent, the highest among 11 Asian economies surveyed (Ronda, 2023). To address these challenges, it is crucial to implement a robust bot-detection system in Shopee's "Product Ratings" to protect users from scams and misleading advertising. This advanced system, capable of analyzing English and Filipino user feedback, will effectively filter out illegitimate reviews. By integrating such a solution, Shopee can ensure the authenticity of product ratings, creating a more transparent and trustworthy online shopping environment for all users. This will not only reduce the risks linked to acquiring defective or falsely advertised products but also actively contribute to solving the problem of identifying between real human reviews and computer-generated information.

To measure the effectiveness of this solution, key performance indicators (KPIs) can be implemented, including a significant reduction in reported scams related to product misrepresentation and an increase in user satisfaction ratings specifically related to product quality and accuracy. These measurable criteria will provide tangible evidence of the positive impact of the bot-detection system on Shopee's user experience and safety.

FISHBONE DIAGRAM:

MANPOWER

Training Knowledge Gap
- Limited exposure to regional dialects and slang hampers context understanding
- Lack of cultural sensitivity training leads to misinterpreting region-specific expressions

Inadequate Language Proficiency
- Reviewers unfamiliar with emojis, abbreviations, impacting human-like expression detection
- Limited awareness of evolving linguistic trends on social media leads to outdated classifications

METHOD

Complex Review Guidelines
- Lack of illustrative examples makes it challenging to apply abstract concepts
- Guidelines not tailored to varying reviewer expertise levels result in inconsistent interpretations

Lack of Streamlined Processes
- Inefficient communication channels cause delays and misunderstandings
- Absence of a feedback loop perpetuates inefficiencies without continuous improvement

MACHINE

Technical Limitations
- Difficulty in processing complex language structures in Tagalog and English

Data Insufficiency
- Limited training data for English and Tagalog

Maintenance Challenges
- Lack of regular updates or maintenance causing software inefficiencies
- Hardware limitations affecting processing capabilities

Incorrect Specification of Linguistic Features
- Lack of detailed linguistic attributes specified in the training data
- Errors in labeling or annotating data that impacts model accuracy

Improper Storage/Handling of Language Components
- Insufficient storage practices causing data corruption or loss
- Mishandling of linguistic components affecting model training quality

MATERIAL

Environmental Impact on Data Collection
- Data collection disruptions due to natural events (e.g., internet outages, server failures during storms)
- Environmental influences on user review behavior (e.g., seasonal trends)

Uncontrollable Environmental Factors Affecting Review Patterns
- Disruptions in review patterns caused by environmental factors (e.g., power outages affecting user activity)

ENVIRONMENT

Inconsistent Evaluation Criteria
- Ambiguous definitions of terms like "natural language" lead to subjective interpretations
- Varying weights assigned to features cause inconsistencies in the final decision-making process

Data Quality Issues
- Incomplete user profiles affect demographic-based classification accuracy
- Lack of continuous monitoring for data drift results in outdated models and decreased accuracy

MEASUREMENT

Complexities in Accurately Classifying Bot-Generated and Human-Generated Reviews on Shopee within English and Tagalog Language

PROJECT DESCRIPTION:

As technology advances, it brings about significant challenges, and one notable concern is the proliferation of computer-generated reviews on online platforms. This project aims to resolve this issue by detecting the authenticity of reviews on Shopee, an online shopping app, distinguishing between those created by human users and those artificially generated by bots.

REFERENCES:

Ronda, R. (2023, November 24). Survey: Philippines has highest e-shopping scam rate in Asia. The Philippine Star. https://www.philstar.com/headlines/2023/11/24/2313809/survey-philippines-has-highest-e-shopping-scam-rate-asia

Royandoyan, R. (2023, December 18). Philippines emerges as Asia's epicenter for online shopping scams.Nikkei Asia. https://asia.nikkei.com/Business/Retail/Philippines-emerges-as-Asia-s-epicenter-for-online-shopping-scams2

Top eCommerce & Shopping Websites in Philippines Ranking Analysis for November 2023 (2023, December 1). Similarweb. https://www.similarweb.com/top-websites/philippines/e-commerce-and-shopping/