Data Ethics and Anonymization

Joy Machuka - Data Scientist Petstore Kenya

Who we are

★ PetStore is a leading ecommerce company that specializes in supplying everything to do with pet products from food to accessories to medicine and other utilities.

Why Data ethics is important for us:

★ Building trust with customers and regulators.
★ Ensuring compliance with data protection laws (e.g., GDPR, Kenya Data Protection Act).
★ Ethical use of data is central to customer satisfaction and business growth.
★ Eliminate bias when using machine learning and other algorithms.

How PetStore treats the issue of potential identification of specific customers?

We do so using various anonymization techniques:-

1. **Data Masking:** Replacing sensitive data (e.g., names, phone numbers) with pseudonyms or hashed values.
2. **Aggregation:** Grouping data to prevent individual identification (e.g., summarizing sales by region instead of customer).
3. **Differential Privacy:** Adding statistical noise to datasets to obscure individual data points.

Re-identification Risk Mitigation

- Conduct regular risk assessments to evaluate vulnerabilities in anonymized data.
- Use synthetic data for testing machine learning models to avoid real data exposure.
- Implement access controls: Only authorized personnel can access raw datasets.

How do we make sure our data/ML procedures are governed by:

1) Fairness

Avoiding Bias

- We conduct bias audits on datasets to ensure diversity and representativeness (e.g., considering customer location, gender, and income variations).
- We check data distribution and implement methods such as resampling to rebalance data, data augmentation, feature selection(avoiding bias features)
- We implement fairness constraints in ML models to prevent discrimination (e.g., fair treatment for rural vs. urban customers).

Inclusive Policies

- We perform regular feedback loops with diverse customer groups to identify overlooked biases.
- Tailoring product recommendations to align with diverse needs (e.g., medicine accessibility for underrepresented pet species).

How do we make sure our data/ML procedures are governed by:

2) Privacy and Security

Privacy Measures

- We minimize on the data collected and collect only necessary information
- We provide customers with clear, transparent privacy policies.
- We give customers right over their data by allowing them to opt out of data collection and tracking.

Security Controls

- We maintain end-to-end encryption for sensitive transactions and storage.
- We use multi-factor authentication for system access.
- We perform regular penetration testing to identify vulnerabilities.

How do we make sure our data/ML procedures are governed by:

3) Transparency

- We develop interpretable machine learning models to explain how recommendations (e.g., product suggestions) are made.
- We communicate clearly to customers why certain actions (e.g., targeted promotions) are undertaken
- We maintain a customer-facing data ethics policy on the website.
- We publish reports on data usage, including anonymization methods and risk management within the company.

How do we make sure our data/ML procedures are governed by:

4) Accountability

Internal Oversight

- We have a Data governance department that oversees ethical use of data and ML.
- We perform regular audits to ensure compliance with internal policies and external regulations.

Customer Feedback and Redress

- We take feedback seriously there we provide channels for customers to inquire about or challenge decisions made using their data.
- We are committed to correcting issues promptly should ethical concerns arise.

At PetStore Kenya, we prioritize customer trust by embedding fairness, privacy, transparency, and accountability into all data practices.

We look forward to continue partnering with regulators to keep or practices in check and also to develop our ML systems with data ethics consideration.

Any questions ?

Thankyou. 😌