RIVA

Enhancing Demographic Data Quality with AWS

Removing Bias and Promoting Explainability

Raj DasGupta
Chief Technology Officer

Table of **Contents**

Executive **Summary**

In the era of big data, the integrity and fairness of demographic data are crucial for informed decision-making in public policy, resource allocation, and social research. Agencies and research organizations that handle large volumes of demographic data must implement rigorous processes to ensure accuracy and fairness. However, biases can unintentionally seep into datasets, leading to skewed insights or potentially unfair outcomes.

This white paper addresses the challenges in maintaining data quality and mitigating bias in demographic data processing. It explores how Amazon Web Services (AWS) Glue, Amazon SageMaker, and other AWS services can enhance data quality, reduce bias, and improve the explainability of machine learning models applied to demographic data, providing important considerations for your technical approach.

Raj DasGupta
Chief Technology Officer
Raj@RIVASolutionsInc.com

Introduction

Demographic data is vital in shaping public policies, guiding social research, and informing resource allocation decisions. The reliability of this data depends on its quality, influenced by factors like completeness, consistency, and accuracy. However, demographic data is susceptible to various forms of bias, such as sampling, measurement, and prejudicial bias, which can distort insights and raise concerns about fairness and transparency in data-driven decision-making.

The increasing use of machine learning (ML) models to analyze and predict demographic trends highlights the importance of transparency and fairness. Often seen as "black boxes," these models require robust frameworks to ensure their decisions are explainable and unbiased. Today, cloud-native tools like AWS Glue and Amazon SageMaker Clarify offer comprehensive solutions to these challenges, integrating data cleansing, transformation, and bias detection capabilities while promoting explainability and transparency in data processing.

Our Methodology

1. Data Quality Enhancement for Demographic Pipelines

AWS Glue is a fully managed ETL (Extract, Transform, Load) service that plays a pivotal role in improving data quality. It enables the automation of data preparation tasks, including data cleansing and transformation, which are critical for ensuring the accuracy and consistency of demographic datasets. AWS Glue's schema discovery feature automatically detects the structure of datasets, allowing for the normalization and standardization of data across different geographic regions and demographic groups. This is particularly important when processing diverse data sources, as it ensures consistency and reduces the risk of errors during analysis. AWS Glue's data profiling capabilities allow for the identification of anomalies, missing values, and inconsistencies within the data. These issues are flagged during the ETL process and can be addressed through automated transformation scripts, ensuring that the final dataset used for analysis is of high integrity and free from bias-inducing errors. A high-level flow is outlined in Figure 1 below enabling the automatic generation of data quality recommendations. These recommendations can be used to create data quality rulesets, with configurable alerts and actions triggered when data quality issues are detected.

Figure 1.
ETL Process flow using AWS Glue to generate Data Quality rule recommendations

2. Bias Detection and Mitigation in Cloud Native Environments

Bias in data processing and machine learning models can lead to unfair or skewed insights. Amazon SageMaker Clarify is a new service designed to detect and mitigate biases. During the data processing stage, SageMaker Clarify analyzes the dataset to identify potential biases, such as an underrepresentation of demographic groups or disproportionate outcomes for specific segments of the population. This analysis is critical for ensuring the data fed into ML models is fair and representative. In addition to detecting bias, SageMaker Clarify enhances the explainability of ML models by offering insights into feature importance and the impact of various variables on model predictions. This transparency is essential for building trust in ML models, as it allows stakeholders to understand how decisions are made and ensures compliance with regulatory standards. Bias detection mechanisms are outlined in Figure 2 which illustrates the workflow for the bias check, training, tuning, lineage, and model registry.

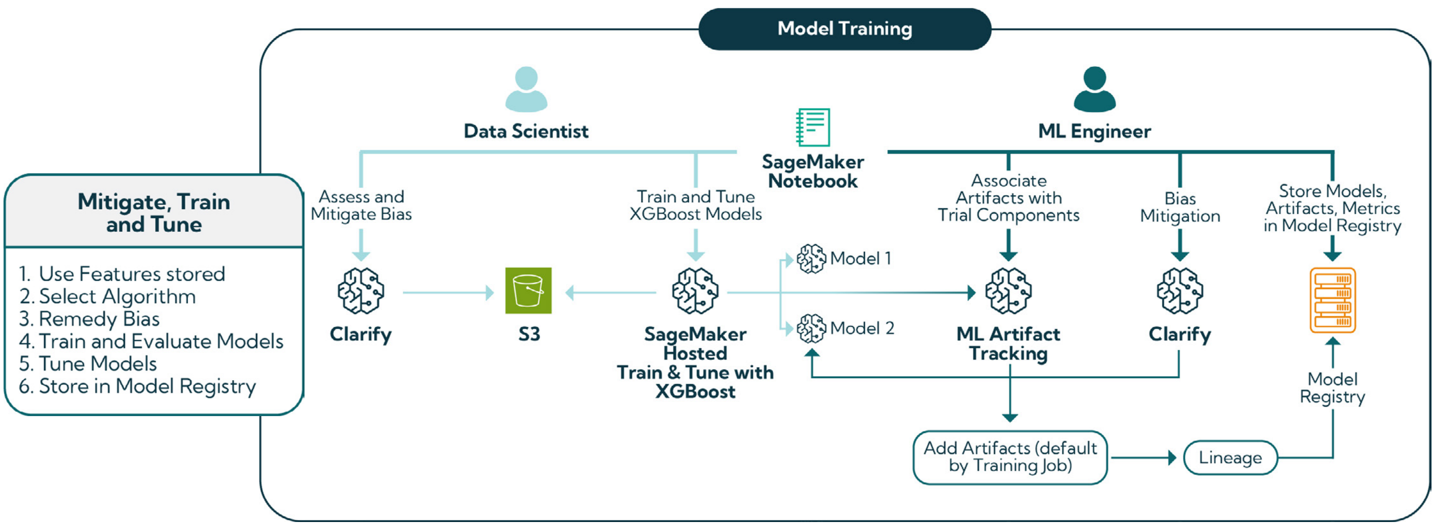

Figure 2.
AWS Sagemaker Clarify architecture for Model Management

3. Continuous Monitoring and Feedback with AWS CloudWatch and SageMaker Model Monitor

Ensuring ongoing fairness and accuracy in models requires continuous monitoring. AWS CloudWatch and SageMaker Model Monitor provide real-time monitoring of data pipelines and machine learning models, allowing organizations to detect and address any issues that arise during production. These tools create alerts for deviations in data processing behavior or model performance, enabling quick interventions to prevent biased data from influencing model outcomes. By integrating monitoring tools into the data pipeline, organizations can establish a robust feedback loop that ensures models remain fair and accurate over time. This continuous monitoring is particularly valuable when dealing with dynamic datasets, such as demographic data, which can evolve and shift over time.

4. Enhancing Demographic Data Processing

Many federal agencies and research organizations process extensive demographic datasets that influence decisions and policy. By utilizing AWS Glue for data preparation, agencies can ensure that all data is accurately cataloged, cleansed, and transformed before analysis. This preparation is crucial for maintaining data integrity and ensuring that insights are based on accurate and representative information. AWS Glue's schema discovery and data profiling capabilities help standardize data across different sources, ensuring consistency and reducing potential bias.

Amazon SageMaker Clarify can then be employed to assess model fairness, ensuring that no demographic group is unfairly impacted by the analysis. Continuous monitoring with AWS CloudWatch and SageMaker Model Monitor ensures that any shifts in data distribution or model performance are detected and addressed promptly, maintaining the integrity of demographic analyses.

Bias can be measured both before and after training a model. SageMaker Clarify provides explanations for model predictions, both post-training and for models deployed in production. For explainability, SageMaker Clarify offers configurations to uncover which features contribute most to model predictions. These results, displayed as SHAP values, reveal the top features with the greatest impact on ML predictions. Results can be viewed in SageMaker Studio or stored in AWS S3 buckets. Additionally, SageMaker Clarify can monitor models in production for any drift in their baseline explanatory attributions and recalculate baselines when needed.

The high-level flow for detecting bias using SageMaker Clarify is illustrated in Figure 3 below.

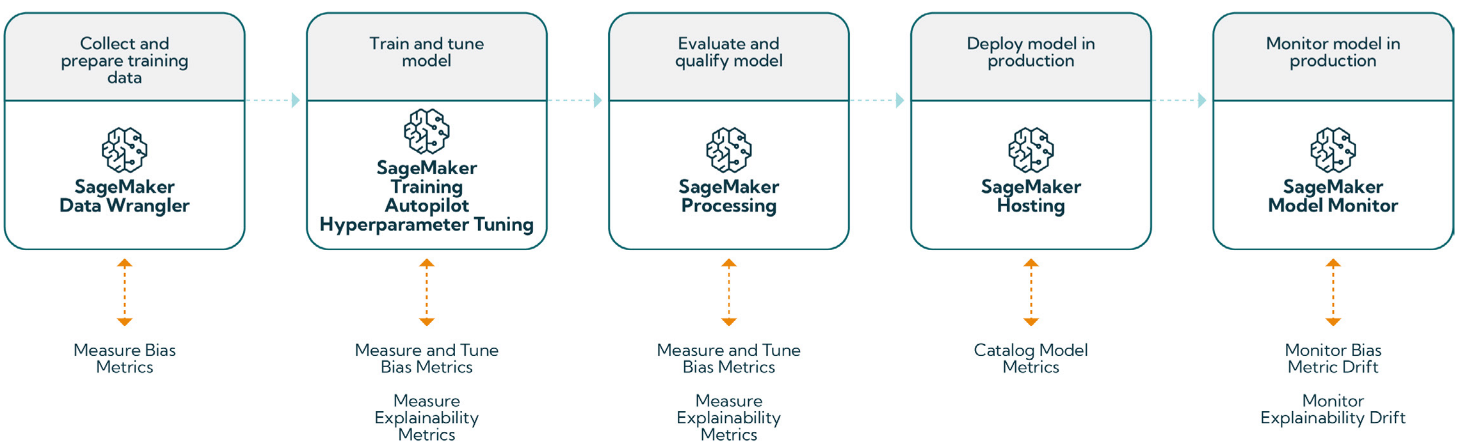

Figure 3.
High Level AWS Sagemaker Clarify process for bias detection

Addressing **Bias**

in Demographic Data Collection and Processing with AWS Tools

To reduce bias in demographic data collection and processing, organizations can leverage various tools and practices within the AWS ecosystem.

Here's how each type of bias can be addressed using this approach:

Sampling Bias

This occurs when the sample used in a study or analysis does not accurately represent the population, leading to skewed results that may not be generalizable to the entire population.

AWS Glue and AWS DataBrew: Use AWS Glue and AWS DataBrew to preprocess and clean your data. You can use these services to identify and correct imbalances in your dataset by ensuring the data is representative of the population. AWS Glue's capabilities can help automate the process of data cataloging and cleansing, making it easier to identify over or underrepresented samples.

Amazon SageMaker Data Wrangler: This service can be used to analyze and visualize data distributions. You can perform statistical checks to ensure your sample data is balanced across different demographic groups before training models.

Amazon SageMaker Ground Truth: Utilize this for creating balanced datasets by defining labeling jobs that ensure all demographic groups are equally represented.

Measurement Bias

Measurement bias happens when the methods or instruments used to collect data consistently distort the data, either overestimating or underestimating the true value, which can lead to inaccurate conclusions.

AWS Lambda and AWS Glue: Implement AWS Lambda functions to validate and standardize measurements before they are ingested into your data lake. AWS Glue can be used to transform data and normalize it correctly, ensuring consistent measurement units and formats across the dataset.

Amazon SageMaker Clarify: This tool can help identify and measure biases in data. By examining feature distributions and performing feature importance analysis, it can help detect and mitigate issues before they affect models.

Exclusion Bias

Exclusion bias arises when certain groups or data points are systematically excluded from the analysis, resulting in a dataset that is not fully representative and may lead to biased outcomes.

AWS Glue: Automate the ETL processes with AWS Glue to ensure all relevant data is included in the dataset. Use Glue's crawlers to automatically detect and include all relevant data sources, reducing the risk of unintentional data exclusion.

Amazon Athena: Use Athena to query your data lake using ANSI SQL to ensure completeness and inclusion of all necessary data. It can help detect potential exclusion by comparing expected versus actual data entries.

Amazon Redshift Spectrum: Analyze large datasets across different sources to ensure no critical data has been excluded, particularly when dealing with massive datasets where exclusion might occur during sampling.

Experimenter or Observer Bias

This bias occurs when the expectations or beliefs of the experimenter or observer unintentionally influence the outcome of the study, often leading to results that reflect the observer's preconceived notions rather than objective reality.

Amazon SageMaker Clarify: This service provides tools to detect potential biases in the data collection process by examining the distribution of data and identifying potential patterns of observer or experimenter bias.

AWS Step Functions: Automate and orchestrate the data collection and preprocessing pipeline using AWS Step Functions to ensure consistency and reduce manual intervention that could introduce bias.

Prejudicial Bias

Prejudicial bias involves the introduction of personal or societal prejudices into the data collection or analysis process, which can lead to unfair or discriminatory outcomes, particularly against certain demographic groups.

Amazon SageMaker Clarify: This service can also be used to detect prejudicial biases in datasets by identifying disparities across different demographic groups. It helps ensure that historical biases do not perpetuate in new models by providing fairness metrics and bias mitigation strategies.

AWS Comprehend: Use AWS Comprehend to analyze text data for biases that might reflect prejudices. This can help in preprocessing stages to filter out or adjust biased content in textual data.

Amazon SageMaker Model Monitor: After deploying models, use this tool to continuously monitor model predictions and ensure that the model is not exhibiting prejudicial bias in production.

Conditional Demographic Disparity (CDD) related Bias

CDD is defined as "the weighted average of demographic disparities for each subgroup, with each subgroup's disparity weighted according to the number of observations it contains." CDD specifically examines how individuals or groups with the same characteristics, apart from the demographic characteristic in question, receive different outcomes. This can indicate bias or unfair treatment in the model.

For example, in a credit scoring model, if two individuals with similar financial profiles but different demographic characteristics (e.g., race, gender) receive different credit scores, this disparity might indicate a CDD. Such bias is particularly concerning because it suggests that the model may be unfairly influencing outcomes based on demographic factors rather than relevant predictors.

Addressing CDD in AWS data pipelines involves a structured approach that spans data preprocessing, model training, and continuous monitoring to identify, measure, and mitigate disparities. This comprehensive strategy leverages various AWS services to ensure fairness throughout the data lifecycle.

Data Preprocessing: AWS Glue can be employed for data cleansing and transformation, using techniques like re-weighting or re-sampling to fairly represent demographic groups. For example, underrepresented groups in a dataset can be oversampled to achieve balance. AWS Glue DataBrew allows for visual inspection and profiling of data, making it easier to identify biases or imbalances. Its recipe feature can normalize or standardize features across demographic groups, reducing the risk of bias entering the model.

Model Training and Post-Processing: Amazon SageMaker Clarify can detect and quantify bias in both data and model predictions by analyzing feature importance and outcome disparities across demographic groups. This helps ensure that models do not inadvertently favor or disadvantage any specific group. Post-processing adjustments, such as applying fairness constraints to model outputs, can be automated with AWS Lambda, while AWS Step Functions ensure consistency across the pipeline.

Continuous Monitoring: Amazon CloudWatch can be configured to monitor pipeline activities and trigger alerts if significant disparities are detected. Regular audits of model performance can be conducted by storing prediction data in AWS S3 and querying it with AWS Athena. Additionally, regularly retraining models using updated and balanced datasets helps keep the pipeline adaptive and responsive to changes, actively reducing CDD and promoting fair outcomes over time.

Best **Practices**

to Evaluate Fairness and Explainability in the ML Lifecycle

Fairness as a Process

The concepts of bias and fairness are context-dependent, influenced by social, legal, and other non-technical considerations. Measuring bias and selecting appropriate bias metrics should be guided by these factors. Successfully adopting fairness-aware ML approaches requires consensus and collaboration among key stakeholders, including product, policy, legal, engineering, AI/ML teams, end users, and communities.

Fairness and Explainability by Design in the ML Lifecycle

It is crucial to consider fairness and explainability at each stage of the ML lifecycle. These stages include problem formulation, dataset construction, algorithm selection, model training, testing, deployment, and ongoing monitoring and feedback. Having the right tools to conduct this analysis is essential. We recommend asking specific questions during each stage of the ML lifecycle, as outlined in Figure 4.

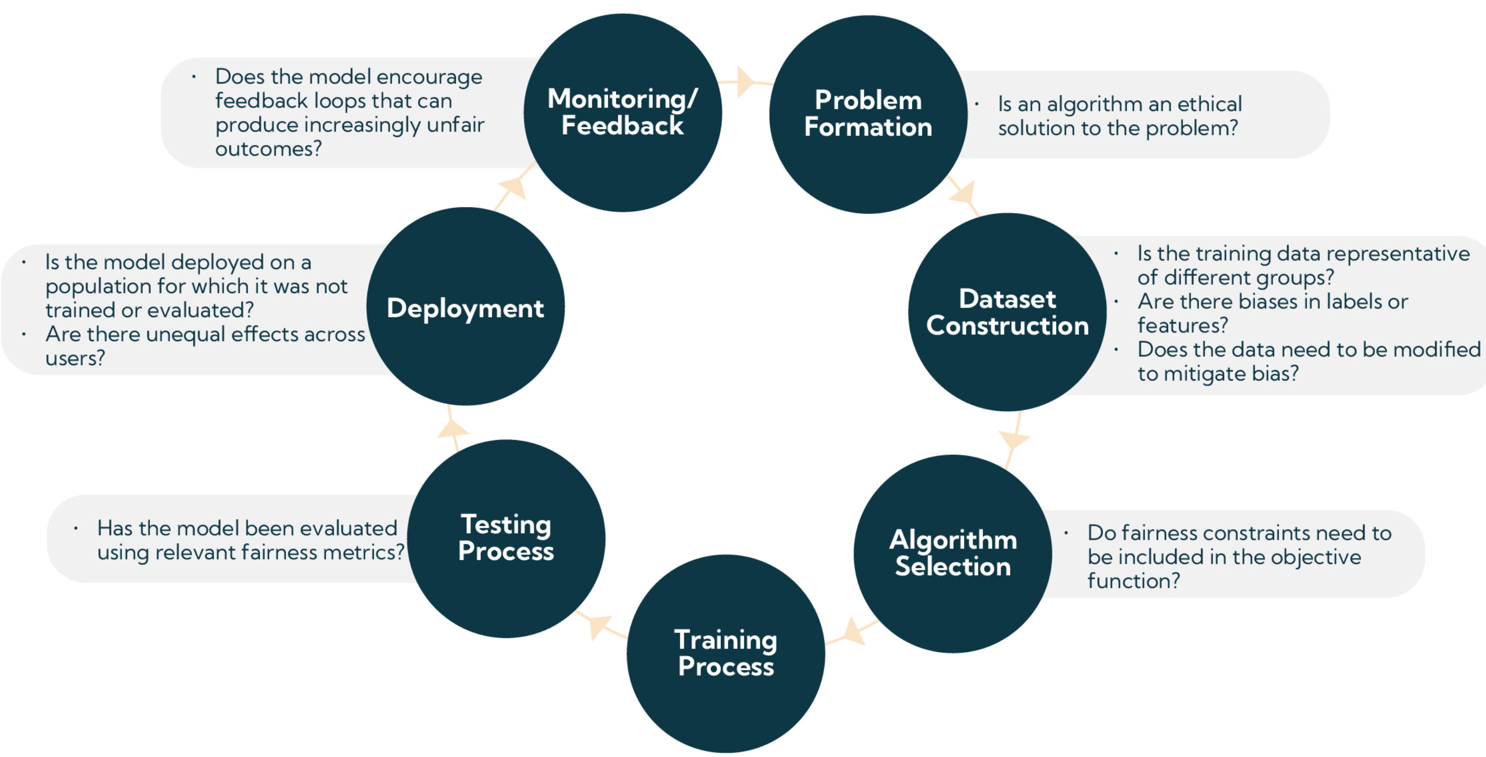

Figure 4.
Fairness and Explainability questions during the ML Lifecycle

Conclusion

AWS Glue, Amazon SageMaker, and other AWS products offer a comprehensive solution for enhancing data quality, reducing bias, and promoting explainability in demographic data processing. By integrating these tools into data pipelines, organizations can build more transparent, fair, and trustworthy models that provide accurate and equitable insights—an essential consideration for organizations like the U.S. Census Bureau, where the accuracy and fairness of data-driven decisions have far-reaching implications.

As the reliance on data-driven insights continues to grow, the need for transparent and fair data processing becomes increasingly critical. AWS provides the tools and infrastructure necessary to meet these challenges, ensuring that demographic data is processed in a manner that is both effective and ethically sound.

Raj DasGupta
Chief Technology Officer
Raj@RIVASolutionsInc.com