

Describing Patterns of Socio-Economic Development at Fine Spatial and Temporal Resolutions

Presented by: Chahat Bansal

Department: School of Information Technology, IIT-Delhi

Young Researcher Symposium, CoDS-COMAD'20

6th January 2020

Socio- Economic Development



Socio-economic development incorporates public concerns in **developing social policy and economic initiatives**

The Indian Census

The Indian Census is among the most credible sources of information on the socio-cultural and demographic data since 1872

The Indian Census

The Indian Census is among the most credible sources of information on the socio-cultural and demographic data since 1872

This is one of the source of primary data at the village, town, and ward level collected once every 10 years

The Indian Census

The Indian Census is among the most credible sources of information on the socio-cultural and demographic data since 1872

This is one of the source of primary data at the village, town, and ward level collected once every 10 years

Census is the basis for reviewing the country's progress in the past decade, monitoring the ongoing schemes of the Government and most importantly, use the data to plan for the future

Limitations of census

EXPENSIVE

Limitations of census

EXPENSIVE

TIME CONSUMING

Limitations of census

EXPENSIVE

TIME CONSUMING

INFREQUENT

Predicting socio-economic indicators with **alternate datasets**



Satellite Data [1]

Predicting socio-economic indicators with **alternate datasets**

[1] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028



Satellite Data [1]



Mass Media [2]

Predicting socio-economic indicators with **alternate datasets**

[1] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028

[2] Sunandan Chakraborty, Ashwin Venkataraman, Srikanth Jagabathula, and Lakshminarayanan Subramanian. 2016. Predicting socio-economic indicators using news events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1455–1464



Satellite Data [1]



Mass Media [2]



Social Media [3]

Predicting socio-economic indicators with **alternate datasets**

[1] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028

[2] Sunandan Chakraborty, Ashwin Venkataraman, Srikanth Jagabathula, and Lakshminarayanan Subramanian. 2016. Predicting socio-economic indicators using news events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1455–1464

[3] Alejandro Llorente Pinto, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. 2015. *Social Media Fingerprints of Unemployment*. (2015).



Satellite Data [1]



Mass Media[2]



Social Media [3]



Agricultural Commodity
Prices [4]

Predicting socio-economic indicators with **alternate datasets**

[1] J Vernon Henderson, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American economic review* 102, 2 (2012), 994–1028

[2] Sunandan Chakraborty, Ashwin Venkataraman, Srikanth Jagabathula, and Lakshminarayanan Subramanian. 2016. Predicting socio-economic indicators using news events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1455–1464

[3] Alejandro Llorente Pinto, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. 2015. *Social Media Fingerprints of Unemployment*. (2015).

[4] Moncarz, P., Barone, S., & Descalzi, R. (2018). Shocks to the international prices of agricultural commodities and the effects on welfare and poverty. A simulation of the ex ante long-run effects for Uruguay. *International economics*, 156, 136-155.

Objective-1

we intend to build a system that automatically pools data together from diverse data sources, **detects noteworthy facts and trends from the data**

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Districts with more manufacturing and services industries develop faster

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Districts with more manufacturing and services industries develop faster

Female participation in the workforce has decreased, primarily in marginal employment

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Districts with more manufacturing and services industries develop faster

Female participation in the workforce has decreased, primarily in marginal employment

Non-agricultural districts see the greatest improvement in all indicators

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Districts with more manufacturing and services industries develop faster

Female participation in the workforce has decreased, primarily in marginal employment

Non-agricultural districts see the greatest improvement in all indicators

Households prefer to invest in assets first, followed by investment in other indicators

Some Noteworthy Facts and Trends from Data

Government prioritized electrification and lighting over other indicators that depend upon government support

Districts with more manufacturing and services industries develop faster

Female participation in the workforce has decreased, primarily in marginal employment

Non-agricultural districts see the greatest improvement in all indicators

Households prefer to invest in assets first, followed by investment in other indicators

Discretionary spending by households is closely related to literacy and formal employment and not by social infrastructure provisioning by the government.

What is missing ?



Problem Statement

In the space of socio-economic development, the data-driven decision making by social sector and government organizations can lead to misleading policies due to **missing perspectives about the local context.**

Problem Statement

In the space of socio-economic development, the data-driven decision making by social sector and government organizations can lead to misleading policies due to **missing perspectives about the local context.**

Trends and anomalies spotted from the mining of big-data sources often leads to interesting observations but **making sense of these observations requires bottom-up data collected through qualitative means to provide the necessary context for correct Interpretation.**

Problem Statement

In the space of socio-economic development, the data-driven decision making by social sector and government organizations can lead to misleading policies due to **missing perspectives about the local context.**

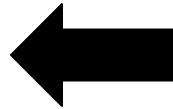
Trends and anomalies spotted from the mining of big-data sources often leads to interesting observations but **making sense of these observations requires top-down data collected through qualitative means to provide the necessary context for correct Interpretation.**

We Propose To Address This Gap

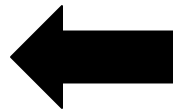
Proposed Technological System



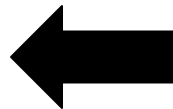
Government



Social Sector



Citizens



Use satellite-data to build ML-based classifiers that make more **accurate predictions of socio-economic indicators** like GDP, HDI, etc.

Validate at higher spatial granularity, several **hypotheses about the relationship between social and economic development**

Qualitative reports can help **provide local context information** which correlates with the anomalous patterns noticed through big-data based observations

Satellite Data

Ground-truth Data

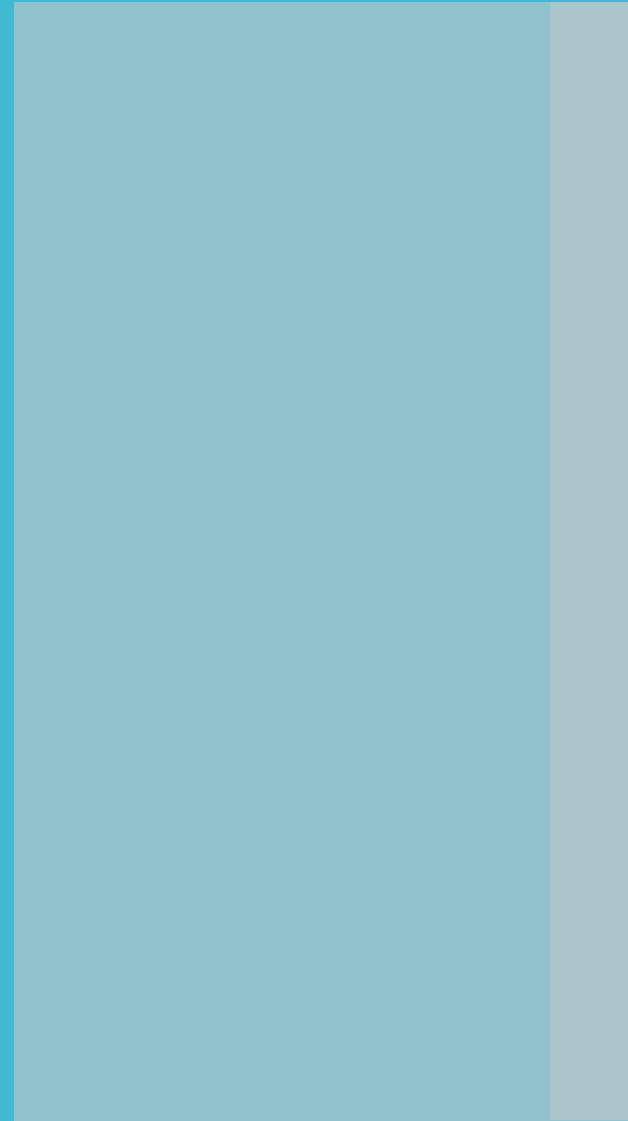
Mass Media/Social Media

Satellite Data

Ground-truth Data

Training Machine Learning Models

Mass Media/Social Media



Satellite Data

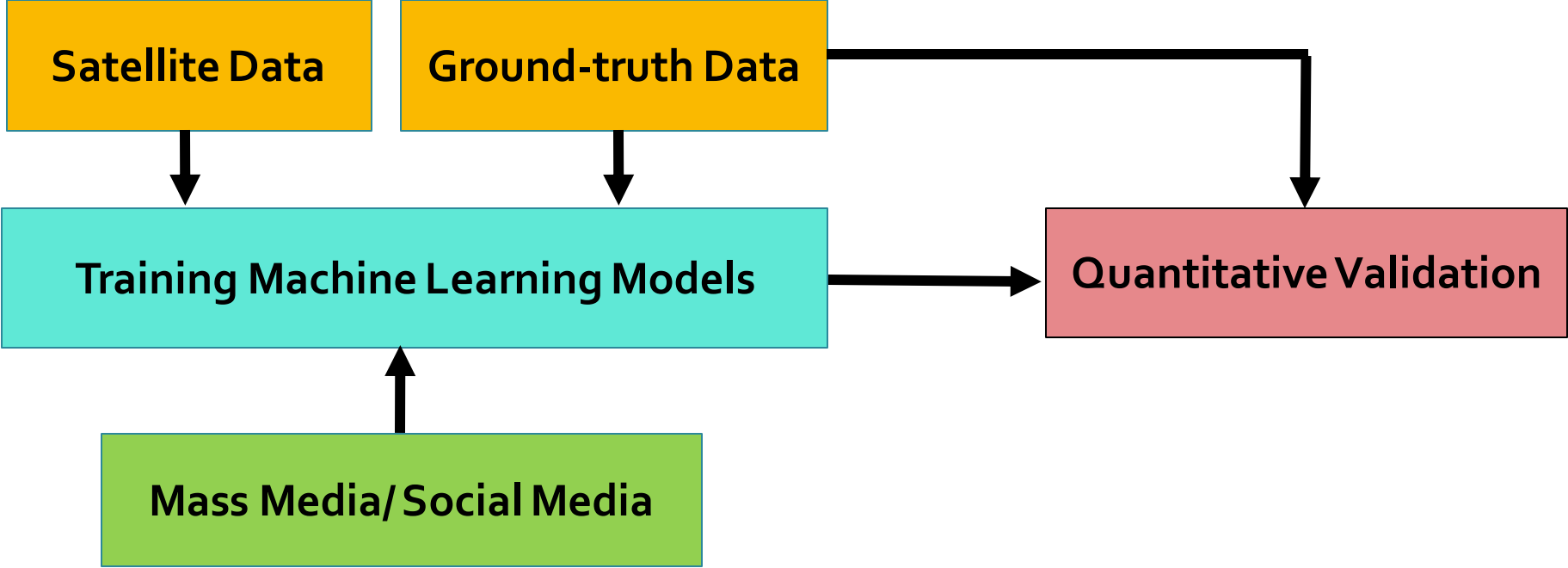
Ground-truth Data

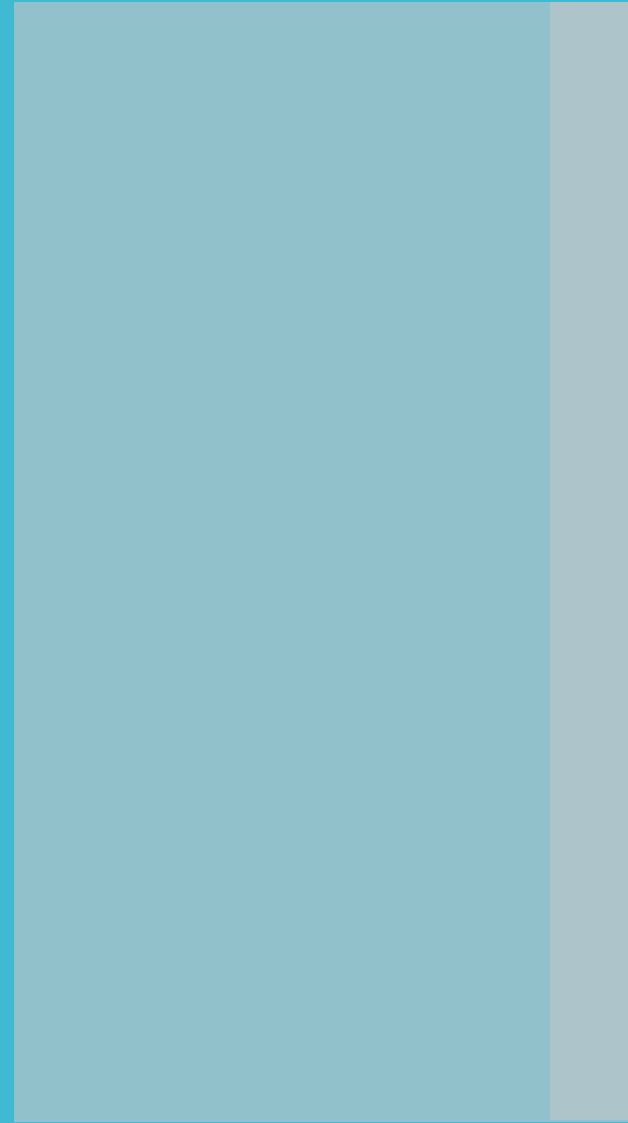
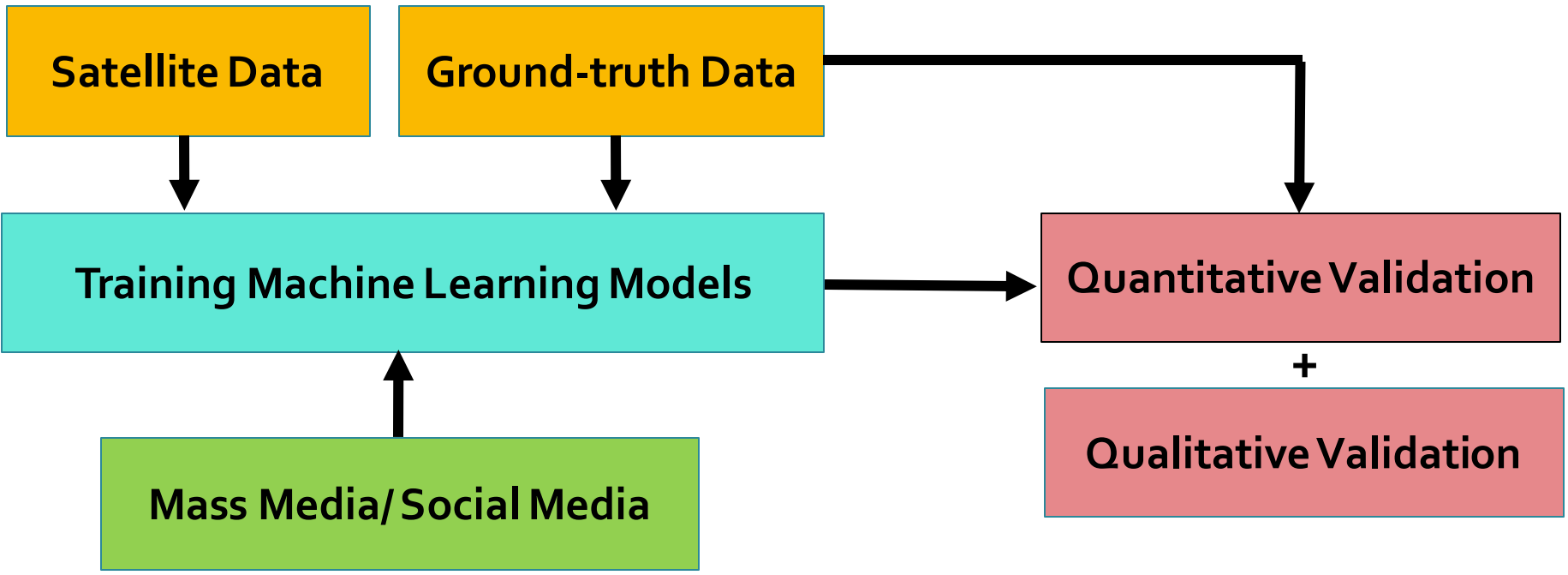
Training Machine Learning Models

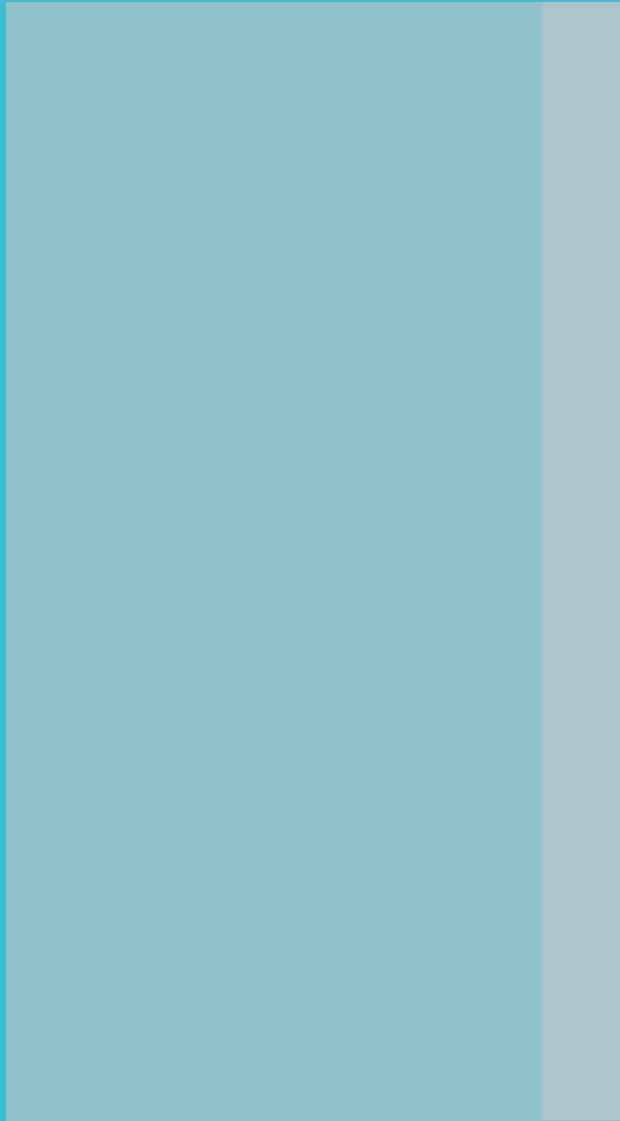
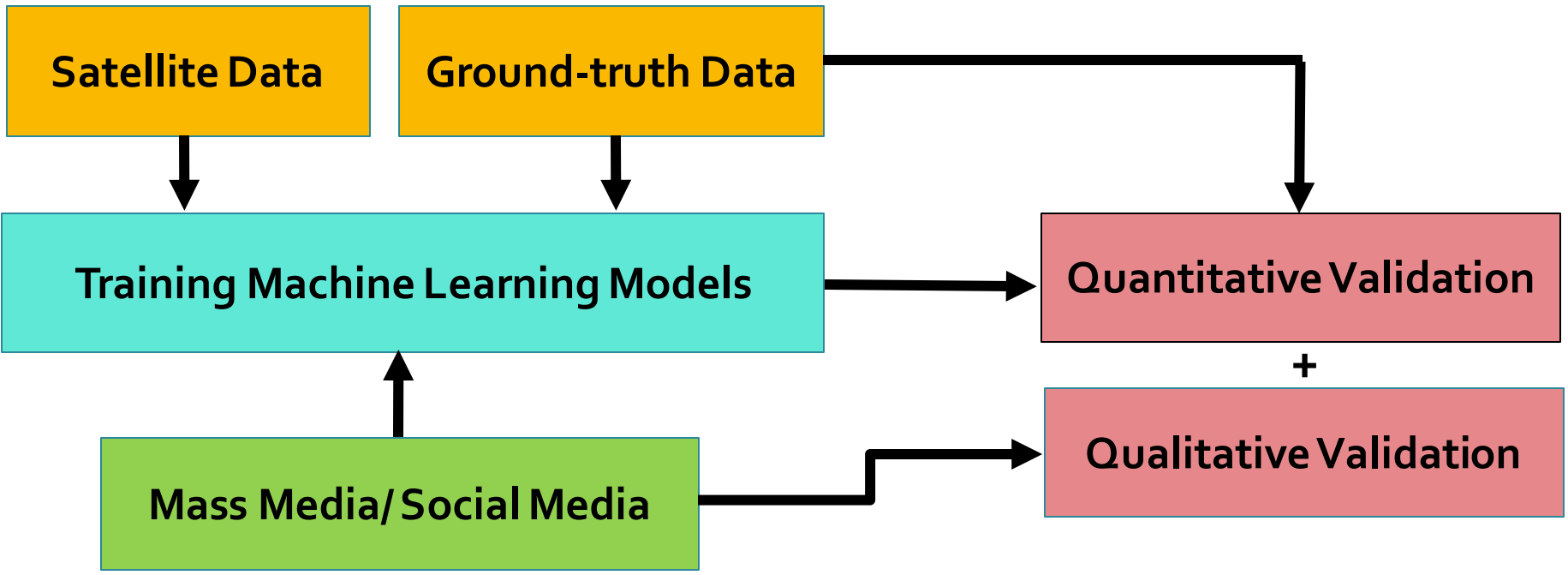
Quantitative Validation

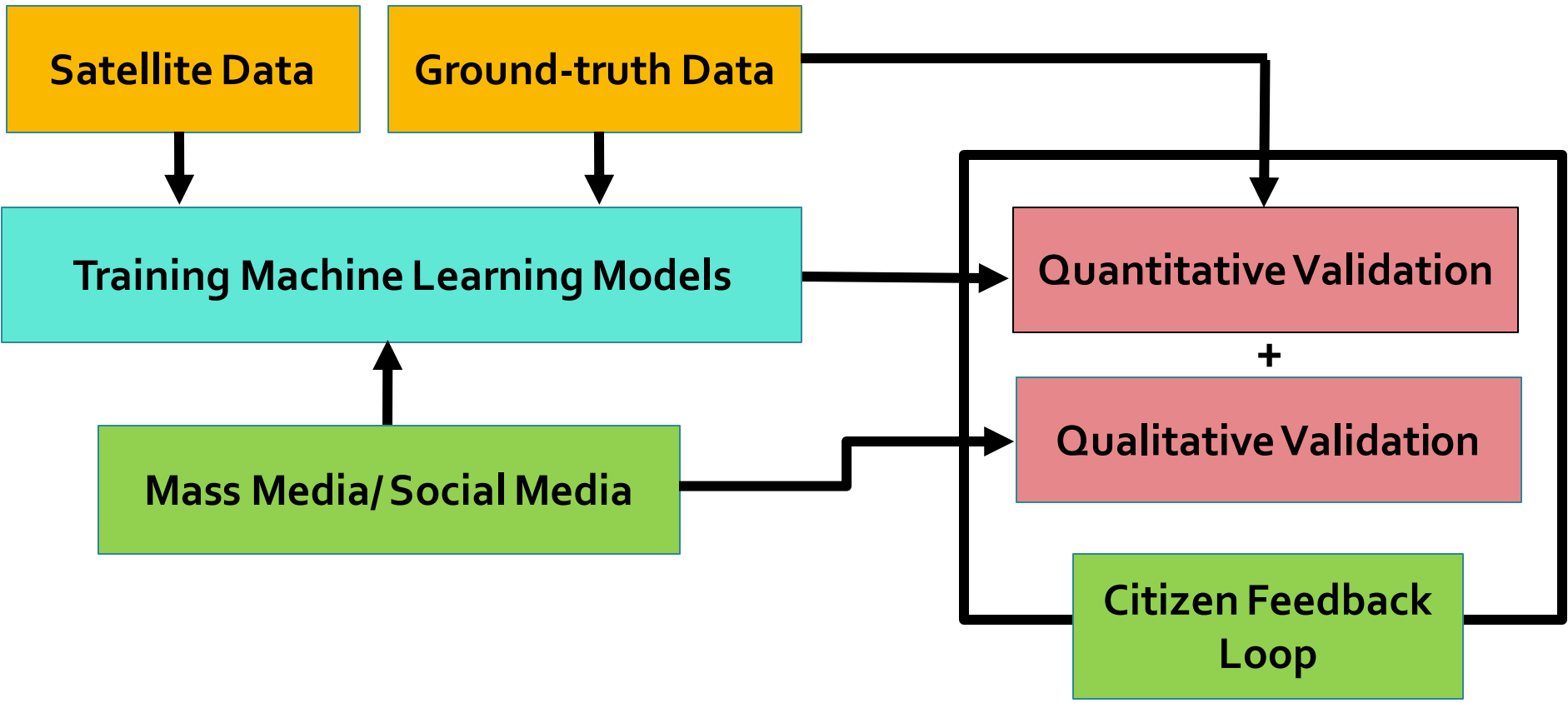
Mass Media/Social Media

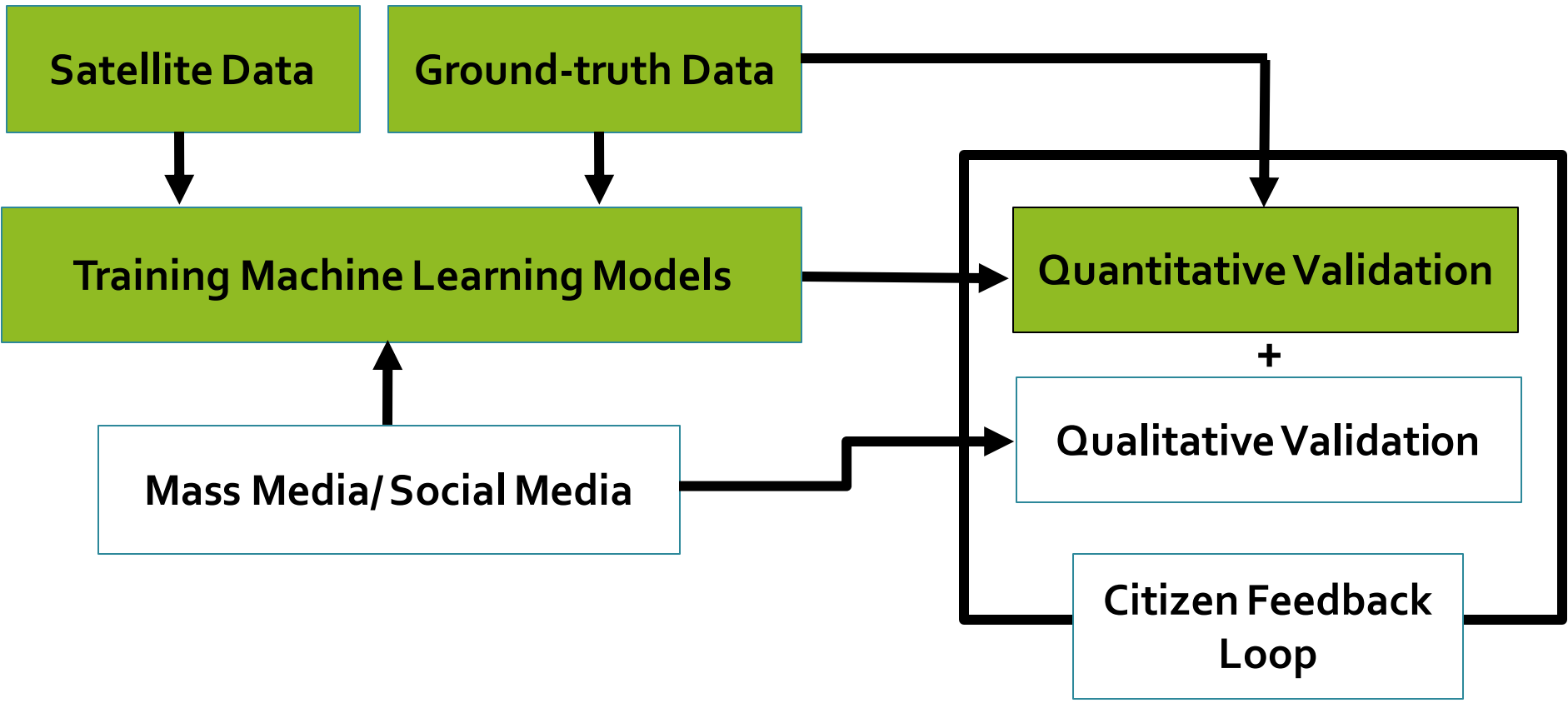




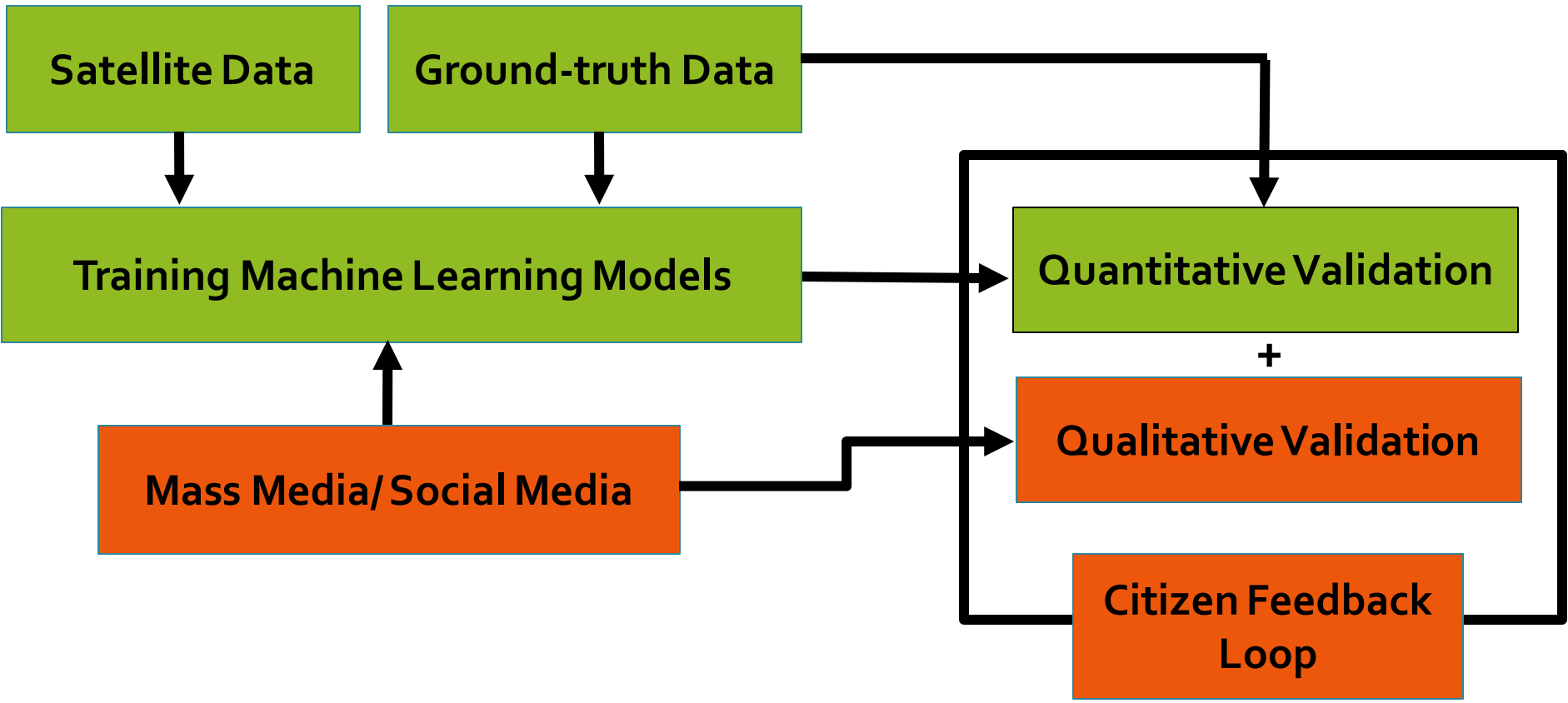










● Work Done



-  Work Done
-  Work In Progress



Conclusion

Potential pitfalls:

- Social media and mass media is susceptible to **bias and misuse**, which can lead to incorrect indicators.
- ML models built on satellite and census data can reflect the **errors and biases** that might be present in the census data.

As part of this venture, we want to specifically explore the potential of three data sources to explain the big-data based observations: **English and vernacular mass media**, **social media**, and **proactively sourced citizen reports** through participatory media platforms.

The image features a white background with two teal-colored geometric shapes. On the left, there is a large teal trapezoid that tapers towards the right. On the right side, there is a smaller teal triangle that tapers towards the left. The text 'Thank You!' is centered between these two shapes.

Thank You!