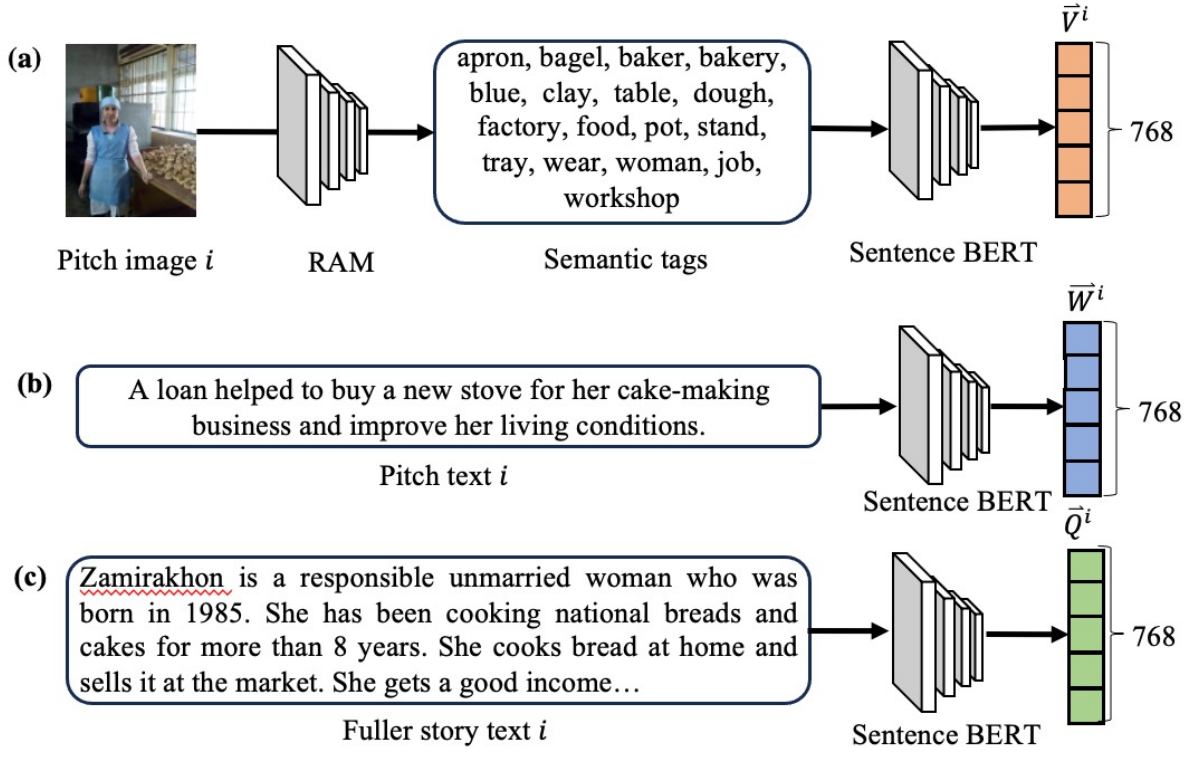
Measurement of Key Variables: An Overview



Let k, j, t be the k-th country, j-th sector, t-th year, respectively, and $\vec{V}^i \in \mathbb{R}^{768}$ be the embedding vector, the cosine distance between image/pitch text/fuller story i and its cluster is

$$\begin{aligned}\text{cosine distance} &= 1 - \cos(\theta) \\ &= 1 - \frac{\vec{V}^i \cdot \text{mean}(\vec{V})}{||\vec{V}^i|| \cdot ||\text{mean}(\vec{V})||} \\ &= 1 - \frac{1}{KJT} \sum_{k,j,t \in K,J,T} \frac{\vec{V}^i \cdot \vec{V}^i_{k,j,t}}{||\vec{V}^i|| \cdot ||\vec{V}^i_{k,j,t}||},\end{aligned}$$

where $\text{mean}(\vec{V})$ is the average vector representations of all other $\vec{V}^i$ in the same country, sector, and year, and K, J, T are the set of countries, sectors, years, respectively.