## $Invalid Schools\_Meta$

14 October 2024 11:39

## Problem Statement:

How many high schools that people have listed on their profiles are real? How do we find out, and deploy at scale, a way of finding invalid schools?

## Solution

<b>→</b>	Algorithm	Random Forest:  • Why Suitable: Random Forests can handle complex patterns and interactions in the data. They are especially useful when multiple features (text features, geographic data, metadata about schools) need to be combined for decision-making. It also handles missing data well and is resistant to overfitting with proper tuning.  • How to Use: In this case, features could include string similarity scores, geographic distances between user and school location, whether the school name appears in known fake school lists, etc. The random forest model would learn to classify schools as valid or invalid based on this data.	
Data Pipelines  Set up automated data pipelines that:  • Regularly scrape and update verified school databases from trusted sources.  • Automate the matching process using string matching and geographic filters in real time.  • Flag potential invalid schools for further review or alert users if they have listed a non-existent school.			VALID SCHOOL INVALID SCHOOL
Data Collection  • Meta School pages (Meta API)  • National Education Council Data (E.g. AICTE, NIRF,etc)  • Education Board Data (E.g. CBSE, SSC, etc)  • Profiles having the same school name  • Location of the school (if mentioned in the profile)			1
	API Integration	The platform sends the school name to the API, which checks it against the database.	