**Project Step 1 - Identifying a Problem (Data Mining Task)**

Anagha Nitin Navale

Harsh Ketan Khona

Jayesh Ramesh Borkar

Stuti Pandey

Information Technology, Arizona State University

IFT 511: Analyzing Big Data

Professor: Asmaa Elbadrawy

January 21, 2024

**Problem Statement:**

The project's aims to develop a model to assist parents identify which schools in a given city are the top five most comparable to a certain school. In order to assess similarity, the model will make use of pertinent school traits and performance indicators. This will give parents insightful information about schools in their chosen city that have comparable features and performance levels. The principal aim of this initiative is to facilitate parents' decision-making process for their children's education by providing a list of carefully selected schools in the same city that share many essential characteristics with a particular school.

**Data Mining task: Similarity Matching:**

The use of similarity matching is likely aimed at helping parents find schools that are similar to a given school within a specific city. Similarity matching allows the model to compare various features or characteristics of schools, such as academic performance, facilities, extracurricular activities, or other relevant factors. Choosing a school is a crucial decision, and having a similarity matching model can serve as a decision support tool. It helps parents explore schools that are comparable to their initial choice, giving them a broader perspective before making a decision.. The model likely uses features or characteristics of schools as input, and the similarity metric helps quantify how closely one school resembles another. This could involve techniques such as cosine similarity, Euclidean distance, or other similarity measures depending on the nature of the data.

Overall, the goal is to enhance the school selection process for parents by leveraging a similarity matching model to provide relevant and personalized recommendations.

**Benefits of solving the problem:** Some of the major benefits of solving this problem are listed below:

* **Improved school comparison:** Parents can learn more about a school's real nature by looking beyond its simple rankings or test results. The algorithm can consider various factors like academic focus, extracurricular activities, student demographics, and learning environment, providing a meaningful comparison.
* **Finding the right one:** Parents may fail to notice excellent schools that are not part of the conventional "top" tier or that are not tailored to their particular needs. Schools that share beliefs and priorities that are in line with their child's unique requirements and learning style can be found through similarity matching.
* **Reduced search time and effort:** Selecting the best school can be difficult and time-consuming, necessitating visits to the institution. By streamlining the process and providing a selection of schools that might be good fits, our solution helps parents save time and money.
* **Data-driven decision making:** The approach, which makes use of sophisticated analytics and similarity matching algorithms, offers a data-driven method for choosing schools. This strengthens the decision support by improving the recommendations' correctness and dependability. This more comprehensive knowledge can help them make better decisions and help their child attend a more rewarding school.

In conclusion, resolving this data mining issue promotes informed choices, a diversity of educational experiences, and community involvement in the local educational system, all of which benefit individual parents in their school selection process and enhance the general empowerment and well-being of the Phoenix metropolitan community.

**The collection of data elements for the selected problem statement:**

Our team aims to build a school similarity matching model that will allow parents to find the most comparable schools to any given school within a specific city. We have identified key attributes like academics, extracurricular activities, demographics, and resources that characterize a school's environment. By evaluating schools based on relevant features beyond just enrollment size or location, our model will provide parents with data-driven school recommendations.

* School Name - Uniquely identifies each school.
* City - Schools should be in the same metropolitan area for an accurate comparison.
* Student Enrollments - Total number of students attending the school.
* AP Classes - Availability of advanced high school courses that can earn college credit.
* Dual Enrollment - Opportunities to take college courses while still in high school.
* Sports Offered - Breadth of competitive, organized athletic programs.
* Teaching Method - Overarching educational philosophy that shapes instruction.
* Math Scoring - Performance on standardized tests for mathematics.
* English Scoring - Performance on standardized tests for English language arts.
* Student-Teacher Ratio - Number of students per instructor in the classroom.
* Racial Diversity - Representation of different races and ethnicities in the student body.
* Free/Reduced Lunch - Percentage of students eligible for meal assistance based on financial need.