Brandon Mather

Project Proposal

Predictive Modeling for High School Student Behavior Trends

Topic:

This project aims to predict future trends in high school student behaviors and interests based on historical data from social networking profiles, spanning the years 2006 to 2009.

Business Problem:

The project addresses the need to understand and anticipate shifts in teenager behaviors and interest, which can be valuable for businesses targeting this demographic, educators designing curriculum, and policymakers who want to help address youth needs more effectively. Key research questions would be:

- Can past trends in online activity and interests among high school students predict future trends?
- Which activities and interests are likely to become popular among high school students in future years?

Datasets:

The dataset I will be using is from Kaggle. It's a random sample of 15,000 high school student's profiles from social networks, spanning the years 2006 to 2009. The dataset was obtained by using text mining techniques. It includes student demographic information and the counts of the 37 most dominant words found in the profiles.

Methods:

The methods that I will use will include:

- Data Cleaning – Handling missing values and feature engineering.
- Exploratory Data Analysis – Understand and identify trends and correlations.
- Predictive Modeling - Use regression or classification algorithms to forecast future trends.
- Model Evaluation – Evaluate model performance using metrics such as accuracy, precision, recall, or RMSE.
- Analysis – Analyze model results.

Ethical Considerations:

Ethical considerations would include making sure the privacy of individuals and sensitive information is handled responsibly, and to be mindful of any biases in the data and model predictions.

Challenges/Issues:

Challenges will include handling missing values and any inconsistencies that exist in the dataset, making sure the predictive model is easy to understand and transparent, and understanding that historical data has limitations for predicting future trends.

References:

To validate my results, I can use case studies and reports from companies that have performed similar predictive modeling projects on teen behavior. For example, the Pew Research Center has done many reports on insights into online behaviors and interests among teens.