Implementing the Teacher–Student Model for Predicting HPV Prevalence

Sainath R

Introduction to Teacher–Student Model

The **Teacher–Student model**, also referred to as **Knowledge Distillation**, is a machine learning framework in which a large, complex, and high-performing model (teacher) transfers knowledge to a smaller and simpler model (student).

The objective is to retain predictive performance while achieving:

- Reduced computational complexity
- Lower memory usage
- Faster inference

Key Criteria for Teacher–Student Modeling

Teacher Model Quality

- Must be well-trained with strong predictive performance.
- Should capture complex relationships in the data.

Student Model Design

- Typically smaller and computationally efficient.
- Should be expressive enough to learn distilled knowledge.

Soft Targets for Distillation

- Uses probabilistic outputs from the teacher.
- Preserves dark knowledge beyond hard labels.

Student Model Features (Part I)

- State
- Patients Attended
- Diabetes
- Hypertension
- HIV (2018–19)
- HDI Score (2017–18)
- Life Expectancy (SRS 2021–25, Male)
- Life Expectancy (SRS 2021–25, Female)
- Marital Age (2017, All Ages)
- Mean Schooling (Rural Male)
- Mean Schooling (Rural Female)
- Mean Schooling (Rural Person)
- Mean Schooling (Urban Male)
- Mean Schooling (Urban Female)

Student Model Features (Part II)

- Population (Females)
- TB Population (2021, in Lakhs)
- Anaemia Prevalence among Women (15–49 years)
- Contraception Use (Any Method)
- Contraception Use (Any Modern Method)
- Female Sterilization
- Male Sterilization
- IUD / PPIUD Usage
- Number of Doctors (2020)
- Expected Years of Schooling (Male)
- Expected Years of Schooling (Female)
- Per Capita Income (2022–23)
- Total Fertility Rate (2018)

Student Model Pipeline

Missing Value Handling

- Median imputation was used due to skewed feature distributions.
- Robust to outliers and non-normality.

Feature Scaling

- StandardScaler was applied to normalize features.
- Ridge Regression is sensitive to feature scale.
- Scaling ensures stable optimization.

Student Model Evaluation Results

Performance Metrics

- Distillation RMSE: **4.2497**
- Distillation R^2: **0.5859**
- Spearman Correlation: **0.7164**

Interpretation

- The student model explains nearly 59% of the teacher model variance.
- Strong rank correlation preserves state-wise HPV prevalence ordering.

State-wise Predicted HPV Prevalence