Report: HVP Rankings Prediction
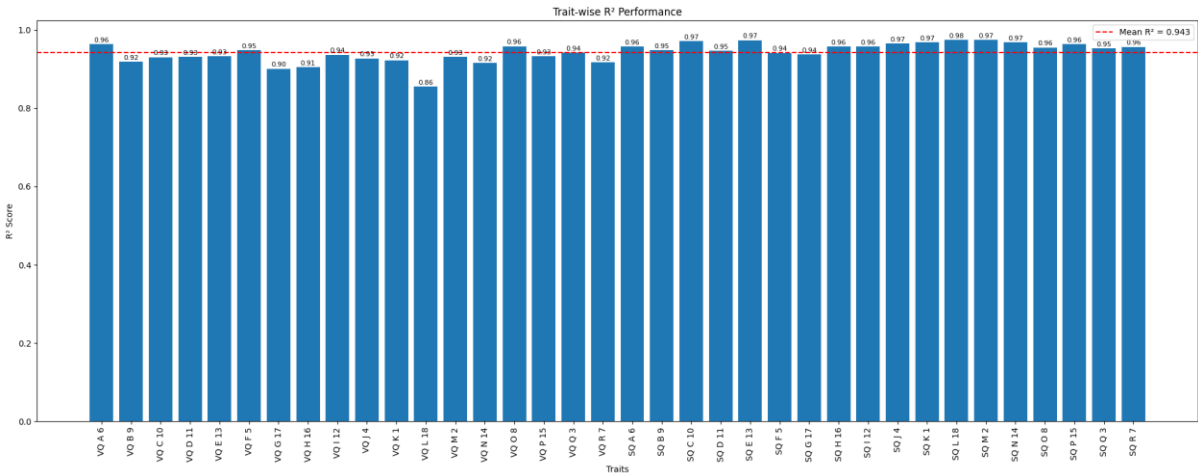
This report summarizes the machine learning pipeline created using SBERT embeddings and an XGBoost multi-output regression model. The goal of the project was to predict 36 personality ranking scores from patient text data.
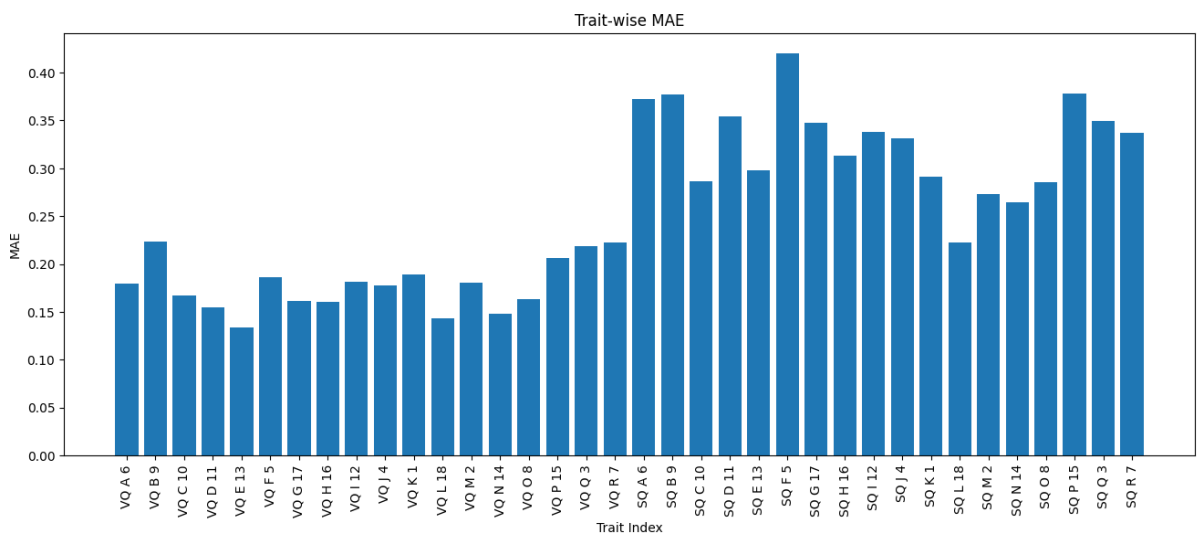
Model Summary

1. Text Embedding: **Sentence-BERT (SBERT)**
2. Encoder Model: **all-mpnet-base-v2**
3. Embedding Dimension: **768**
4. Regression Model: **XGBoost (Multioutput Regressor)**
5. Achieved R^2 Score: **0.94**

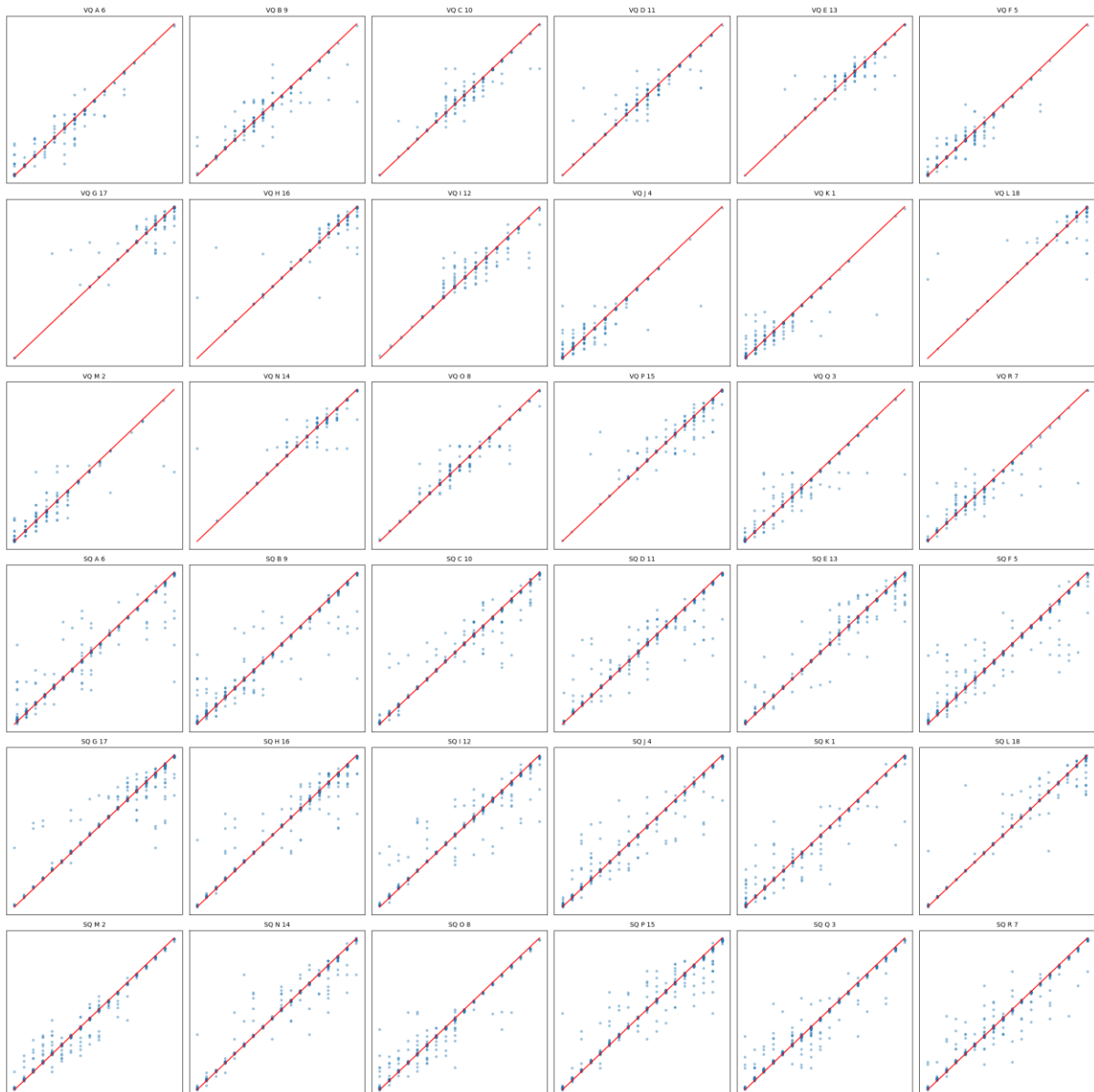**Trait wise R^2 Bar Chart**

MAE Bar Chart

True vs Predicted Scatter Plot Grid (36 Rankings)

Residual Plot Grid (36 Rankings)

VQ A 6	VQ B 9	VQ C 10	VQ D 11	VQ E 13	VQ F 5
VQ G 17	VQ H 16	VQ I 12	VQ J 4	VQ K 1	VQ L 18
VQ M 2	VQ N 14	VQ O 8	VQ P 15	VQ Q 3	VQ R 7
SQ A 6	SQ B 9	SQ C 10	SQ D 11	SQ E 13	SQ F 5
SQ G 17	SQ H 16	SQ I 12	SQ J 4	SQ K 1	SQ L 18
SQ M 2	SQ N 14	SQ O 8	SQ P 15	SQ Q 3	SQ R 7