**ARTIFICIAL INTELLIGENCE OPTIMIZATION FOR COMPUTER SCIENTISTS — D682**

TASK 3: AI MODEL OPTIMIZATION

**A1. Distribution of Three Key Variables**

I examined **healthRiskScore**, **tempmax**, and **humidity** across 1,000 observations. Summary statistics indicate:

* **Health Risk Score:** Mean = 9.73; median = 9.55; range = 8.49–11.49; right‑skewed with tail of higher‑risk days.
* **Maximum Temperature (tempmax):** Mean = 85.1 °F; median = 84.3 °F; range = 62.04–107.8 °F; approximately normal distribution, with extremes at < 70 °F and > 100 °F.
* **Relative Humidity:** Mean = 56.8 %; median = 58.5 %; range = 11.75–92.46 %; spread across mid‑range values.

Notable Trend: Occasional spikes in healthRiskScore appear associated with extreme weather conditions (very high temperature or humidity), warranting deeper pattern analysis.

**A2. Underlying Patterns**

1. **Diurnal Cycle of Health Risk**
   * Observation: Average healthRiskScore peaks at **5 AM (10.02)** and **6 PM (9.99)**, with a minimum around **4 AM (9.54)**.
   * Evidence: Pattern extracted from patterns\_hourly\_risk.csv with grouped hourly means.
2. **Meteorological Correlations**
   * Observation: healthRiskScore exhibits moderate positive correlation with tempmax (r = 0.35) and weaker with humidity (r = 0.19); temperature and humidity are strongly negatively correlated (r = –0.61).
   * Evidence: Correlation matrix from patterns\_correlation.csv.

**A3. Hypothesis**

The early‑morning (5 AM) and early‑evening (6 PM) peaks in healthRiskScore are driven by rapid temperature increases following overnight low‑humidity conditions, which cause pollutant accumulation in stagnant air and lead to elevated health‑risk measurements.

**B1. Performance Metrics Analysis**

I evaluated four models using **RMSE** and **MAE**:

* **Baseline GBM:** RMSE 0.196; MAE 0.153.
* **Optimized GBM:** RMSE 0.195; MAE 0.152 (0.5 % improvement).
* **Stacking Ensemble:** RMSE 0.120; MAE 0.096 (38.8 % reduction in RMSE).
* **Voting Ensemble:** RMSE 0.131; MAE 0.106 (33.2 % reduction in RMSE).

Lower RMSE and MAE indicate tighter prediction intervals and improved reliability in forecasting healthRiskScore, enabling more accurate alerts and resource allocation.

**B2. Insights from Model Behavior**

Feature importance from the optimized GBM reveals:

1. **dew (≈ 35.5 %)**: Highest driver, highlighting moisture’s critical role in health risk.
2. **heatIndex (≈ 23.3 %)**: Significant contributor, underscoring perceived temperature impact.
3. **severityScore (≈ 12.2 %)**: Reflects influence of past event severity on current risk.

These insights suggest focusing monitoring and mitigation efforts on dew‑point and heat‑index conditions.

**B3. Future Predictions & Strategic Insights**

1. Use **dew point forecasts** to trigger proactive health‑risk alerts during high‑moisture periods.
2. Leverage **heatIndex thresholds** in operational models to adjust alert criteria dynamically based on perceived temperature.
3. Integrate **severityScore trends** into monitoring dashboards for scheduling maintenance and sensor calibration.