# 📄 INTEGRATED DISEASE SURVEILLANCE & PREDICTIVE VACCINATION STRATEGY

## 1️⃣ Project Vision

To establish a robust **Integrated Disease Surveillance System (IDSS)** that empowers health departments to:

* Monitor outbreaks *in near real-time* at the district level.
* Identify *root causes* such as vaccination gaps and weather-driven factors.
* Predict *next week’s likely high-risk districts* with actionable ML insights.
* Enable targeted vaccination drives, rapid containment, and optimized resource deployment.

## 2️⃣ Strategic Objectives

✔️ Detect emerging disease hotspots early.  
✔️ Close vaccination gaps proactively.  
✔️ Use weather-linked insights to forecast spikes.  
✔️ Enable district collectors and health officers with weekly ready-to-use reports.  
✔️ Build trust with the public through transparent, data-driven interventions.

## 3️⃣ Data Sources & Coverage

This surveillance system integrates **multi-source datasets**, including:

* Weekly outbreak reports from IDSP.
* District-level vaccination coverage.
* Official population estimates.
* District-wise weather data (rainfall & humidity).

📍 **Scope:**

* Covers all districts with unique district\_code and district\_name.
* Merges historical & real-time feeds.
* Fully scalable to new diseases, regions & future health missions.

## 4️⃣ End-to-End Execution Plan

### 🔹 PHASE 1 — Data Cleaning & Master Dataset

* Collected raw cases, vaccination, population & weather data.
* Merged into a **single master file**.
* Added derived metrics: *Incidence Rate (cases per 1,000)*, *Risk Flag* (based on threshold).

### 🔹 PHASE 2 — Exploratory Data Analysis (EDA) & Advanced Statistics

* Plotted **weekly trends**, **district heatmaps**, **pairplots**, **3D weather-case plots**.
* Ran **T-tests, Chi-Square, ANOVA** to prove vaccination’s impact on outbreak control.
* Calculated **correlation** between weather factors (rainfall, humidity) & case spikes.

### 🔹 PHASE 3 — Machine Learning & Forecasting

* ✅ **Logistic Regression Baseline:** Trained a simple, interpretable model to classify districts as high or low risk using lag features, vaccination, and weather data.
* ✅ **Random Forest Classifier:** Added to capture non-linear effects & boost accuracy. Feature importance highlights top drivers (rainfall, past cases, vaccination).
* ✅ **Prophet Forecast:** Built district-level time-series forecasts to project outbreaks *4 weeks ahead*.
* ✅ **Actionable Output:** Risk predictions + forecasts merged into **one risk\_forecast\_output.csv** for real-time decision support.

### 🔹 PHASE 4 — Power BI Dashboard

A 3-page, interactive, DAX-rich, policy-ready dashboard:  
1️⃣ **Outbreak Monitoring:** Weekly trends, heatmaps, top risk districts.  
2️⃣ **Vaccination & Weather Impact:** Choropleth maps, rainfall/humidity scatterplots, low vaccination bars.  
3️⃣ **Predictive Risk & Action Plan:** Forecasted risk map, high-risk flags, recommended next steps.

### 🔹 PHASE 5 — Excel Reports & Pivots

* Final **Integrated\_Disease\_Surveillance\_Report.xlsx** with raw master data.
* Ready-to-slice pivot tables: *District vs Week*, *Incidence vs Vaccination*, *Weather vs Cases*.
* Print-friendly summaries for **offline briefings**.

## 5️⃣ Deliverables

| **✅ Deliverable** | **📌 Description** |
| --- | --- |
| Clean Master CSV | Final structured dataset |
| Python Notebooks | EDA + ML with beginner-friendly 📌 Insight cells |
| ML Output CSV | risk\_forecast\_output.csv — next week’s high-risk districts |
| Power BI PBIX | DAX measures, visuals, map plots |
| Excel Workbook | Pivot tables, charts, summary sheets |
| Execution Plan | [Attached Excel execution plan] |
| Final Report | [This document] |

## 6️⃣ Expected Impact

✔️ Real-time identification of emerging hotspots.  
✔️ Proactive vaccination drives in under-covered districts.  
✔️ Forecast-based early warning for weather-driven outbreaks.  
✔️ Strong evidence base for District Health Officers to justify urgent interventions.  
✔️ Fully replicable for other states & diseases.

## 7️⃣ Next Steps & Recommendations

🔹 Pilot in **high-burden districts**.  
🔹 Connect to **live IDSP feeds & weather APIs**.  
🔹 Automate updates with **weekly ML runs & scheduled refreshes**.  
🔹 Train district-level staff on Power BI usage.  
🔹 Roll out phased scale-up to state & national dashboards.