# **Week 7 Report – Analytics Modeling Phase**

**Week Focus: Data Analytics & Machine Learning Modeling**

During the seventh week of the InsightNation project, the team transitioned into the **Analytics Modeling** phase, a crucial milestone focused on extracting meaningful insights from the cleaned and integrated citizen feedback data collected in earlier weeks. This phase aimed to apply a combination of **data analytics and machine learning (ML)** techniques to discover patterns, identify trends, and support evidence-based recommendations for improving public services.

**1. Data Review and Preparation**

The week began with a comprehensive **review of the final cleaned dataset**, which includes over 4,000 records with feedback across various services like public toilets, transport, parks, libraries, and local governance. Data was rechecked for consistency, normalized across categories, and encoded where necessary for modeling. Numerical and categorical variables were prepared in formats suitable for analytics, ensuring compatibility with ML algorithms.

**2. Selection of Modeling Techniques**

Based on the nature of the feedback—both structured (e.g., satisfaction scores) and unstructured (e.g., free-text suggestions)—appropriate **analytical methods** were selected. These included:

* **Sentiment analysis** on open-text responses using NLP techniques (spaCy + custom rules).
* **Logistic Regression and SVM classifiers** to predict satisfaction levels based on service attributes.
* **K-means clustering** to segment citizens into groups based on usage patterns and opinions.
* **Descriptive statistics and correlation analysis** to understand feature relationships.

The goal was to extract both quantitative and qualitative insights that can later feed into actionable dashboard components.

**3. Model Training and Validation**

Data was split into **training (80%) and validation (20%) sets**. ML models were trained on labeled data to predict satisfaction or classify user sentiment. For example, logistic regression was used to classify sentiment in feedback regarding park cleanliness and transport safety. Performance was measured using accuracy, precision, recall, and F1-score. The team also validated models using **cross-validation techniques** to prevent overfitting and ensure robustness.

**4. Iterative Refinement and Tuning**

After initial results, the models underwent iterative improvements. Hyperparameter tuning was applied using **GridSearchCV** for models like SVM to boost accuracy. Stop words were fine-tuned in NLP models, and TF-IDF vectorization was applied to textual features to enhance context detection in sentiment classification. The team also experimented with ensemble techniques like **Random Forest** for higher accuracy in multi-feature predictions.

**5. Insight Documentation**

All modeling outcomes and observed trends were documented in detail. For instance, data showed that citizens aged 25–40 reported the most dissatisfaction with public transport safety, while elderly groups valued clean park environments more. Sentiment analysis of suggestions revealed recurring themes such as **“lighting,” “safety,” and “maintenance”** across most categories. These findings are essential for strategic planning and will be integrated into the dashboard in upcoming weeks.

**6. Team Collaboration and Feedback**

Throughout the week, collaboration was maintained through shared notebooks, daily stand-ups, and code reviews. Peer feedback helped refine model selection strategies and validate assumptions. Issues like imbalanced data in certain categories (e.g., library feedback) were addressed through resampling techniques.

**7. Preparation for Dashboard Integration**

As the week concluded, a curated list of **key insights and predictions** was compiled for integration into the dashboard in the next phase. These included:

* Most impactful service attributes by category.
* Key complaint themes extracted via NLP.
* Sentiment scores and citizen satisfaction predictions.

By the end of Week 7, the InsightNation project successfully completed the core analytics modeling tasks and generated actionable insights. These insights will serve as the analytical backbone for the dashboard and reporting features planned for the upcoming development phases.