DataAI - Web Analytics AI Agent

Author: Andriy Uspishnyi
Date: June 6, 2025

Introduction

This AI Agent is built to provide comprehensive insights into website performance data (user activity) by combining advanced data analysis, machine learning, and natural language processing. This system can be further adjusted for various use cases (eg. industries, websites, etc.) depending on the input data fields available and types of insights required. Let's explore how this sophisticated AI agent works.

System Architecture

I. Core Components

The system consists of several specialized AI agents working together:

1. **Data Analyzer Agent**
    - Handles raw data processing and statistical analysis
    - Performs time series analysis
    - Generates forecasts using XGBoost
    - Creates visualizations and reports
2. **Interpretation Agent**
    - Uses AWS Bedrock (Claude) for natural language understanding
    - Generates human-readable insights
    - Incorporates industry news context
    - Provides actionable recommendations
3. **News Summary Agent**
    - Monitors industry news using Tavily search
    - Provides real-time market context
    - Focuses on relevant automotive industry updates
4. **Podcast Generator**
    - Creates engaging audio content from insights
    - Uses ElevenLabs for text-to-speech conversion
    - Formats content for optimal listening experience

II. Data Flow

The system processes data through the following workflow:

1. Data Ingestion
    - Accepts CSV files with website analytics data
    - Required fields: pixid, unique_pixel, assigned_domain, assigned_brand, dma, dma_name, date, action, hitCount
2. Data Processing

        - Filters data based on user criteria
        - Performs time series decomposition
        - Calculates key metrics and trends
3. Analysis Pipeline
        - Time series analysis with trend detection
        - Seasonal pattern identification
        - Volatility assessment
        - XGBoost-based forecasting
4. Insight Generation
        - Natural language interpretation of results
        - Integration with industry news
        - Generation of actionable recommendations

Key Features

I. Advanced Analytics

1. **Time Series Analysis**
        - Moving averages (7-day and 30-day)
        - Growth rate calculations
        - Trend detection
        - Seasonality analysis
2. **Forecasting**
        - XGBoost-based predictions
        - Feature engineering for time series
        - 14-day forecast horizon
        - Confidence intervals

II. AI-Powered Insights

1. **Natural Language Interpretation**
        - Automated insight generation
        - Context-aware analysis
        - Industry news integration
        - Actionable recommendations
2. **Market Context**
        - Real-time industry news monitoring
        - Automotive market trends
        - Supply chain insights
        - Dealer operations impact

III. Multi-Format Output

1. **Interactive Dashboard**

- Real-time data visualization
- Filterable metrics
- Custom date ranges
- Domain/brand filtering

2. **Exportable Reports**
    - HTML report generation
    - Key metrics summary
    - Visual charts and graphs
    - Forecast projections
3. **Audio Content**
    - Podcast-style summaries
    - Natural-sounding narration
    - Professional formatting

Technical Implementation

I. Machine Learning Components

1. **XGBoost Model**
    - Features: time-based, lag, and rolling mean
    - Hyperparameter tuning
    - Feature scaling
    - Model validation
2. **Time Series Analysis**
    - Seasonal decomposition
    - Trend analysis
    - Volatility metrics
    - Pattern recognition

II. AI Integration

1. **AWS Bedrock**
    - Claude model for interpretation
    - Natural language generation
    - Context-aware analysis
    - Structured output formatting
2. **Tavily Search** *(credits: Evan Murphy)*
    - Real-time news monitoring
    - Relevant content filtering
    - Multi-angle search
    - Content deduplication

III. Audio Generation

1. **ElevenLabs Integration**
    - Natural voice synthesis (specification)
    - Script formatting
    - Audio quality optimization
    - Professional presentation

Best Practices and Usage

1. Data Preparation
    - Ensure CSV format compliance
    - Validate required fields
    - Check data quality
    - Handle missing values
2. Analysis Configuration
    - Select appropriate date ranges
    - Choose relevant filters
    - Set forecast parameters
3. Output Utilization
    - Review generated insights
    - Export reports as needed
    - Listen to audio summaries
    - Implement recommendations

Conclusion

This AI-powered analytics system represents a sophisticated approach to website data analysis, combining multiple AI agents to provide comprehensive insights and actionable recommendations. The system's modular architecture allows for easy updates and extensions, while its multi-format output ensures insights are accessible in various ways to different stakeholders. The integration of machine learning, natural language processing, and audio generation creates a powerful tool for understanding website performance and making data-driven decisions in the automotive industry context.

APPENDIX:

Agent Flow: Step-by-Step Process

1. Initial Data Upload

When a user uploads a CSV file:

1. Data Validation
    - System checks for required columns (pixid, unique_pixel, assigned_domain, assigned_brand, dma, dma_name, date, action, hitCount)
    - Validates data types and formats
    - Handles missing values and data cleaning
2. Data Analyzer Initialization
    - Creates DataAnalyzer instance with the uploaded DataFrame
    - Converts date column to datetime format
    - Prepares data for analysis

2. User Filter Selection

When user selects filters in the sidebar:

1. Data Filtering Process
    - Applies date range filter
    - Filters by selected domain
    - Filters by selected brand
    - Filters by DMA
    - Filters by action type
    - Updates all visualizations in real-time
2. Time Series Analysis
    - Calculates daily aggregations
    - Computes 7-day and 30-day moving averages
    - Calculates daily and weekly growth rates
    - Updates time series visualizations

3. Analysis Generation

1. Statistical Analysis
    - Calculates key metrics:
        - Total hits
        - Average daily hits
        - Unique domains count
        - Unique DMAs count
    - Generates summary statistics
2. Forecasting Process
    - Prepares features for XGBoost model:
        - Time-based features (day of week, month, year)
        - Lag features (1, 7, 14, 30 days)
        - Rolling mean features (7, 14, 30 days)

- Trains XGBoost model
- Generates 14-day forecast
- Calculates confidence intervals

4. AI-Powered Insights Generation

1. News Summary Agent Activation
    - Searches for relevant industry news using Tavily
    - Performs multi-angle search:
        - Current market conditions
        - Inventory levels
        - Vehicle pricing
        - Supply chain issues
        - EV sales trends
        - Dealer news
    - Deduplicates and filters news results
2. Interpretation Agent Processing

- Prepares data summary for Claude
- Generates natural language insights:
    - Overview of key metrics
    - Trend analysis
    - Seasonal patterns
    - Volatility assessment
    - Forecast interpretation
    - Industry context integration
    - DMA-specific insights
    - Actionable recommendations

5. Output Generation

1. Interactive Dashboard Updates
    - Updates all visualizations:
        - Time series plots
        - Forecast charts
        - Key metrics displays
        - DMA performance charts
    - Refreshes filter options based on available data
2. Report Generation
    - Creates HTML report with:
        - Key metrics section
        - Time series analysis
        - Forecast projections
        - AI-generated insights
        - Industry news context
        - PRIZM recommendations
3. Podcast Generation

- Creates podcast script:
    - Introduction
    - Key findings
    - Market context
    - Recommendations
    - Musical interludes (placeholders)
- Converts script to speech using ElevenLabs
- Adds professional formatting and pauses

6. Export Options

1. HTML Report Export
    - Generates comprehensive HTML report
    - Includes all visualizations
    - Embeds interactive elements
    - Provides downloadable format
2. Audio Export
    - Generates MP3 file of podcast
    - Includes professional formatting
    - Adds musical interludes
    - Optimizes audio quality
3. Data Export
    - Provides filtered dataset
    - Includes forecast results
    - Exports in CSV format
    - Includes metadata

7. Continuous Updates (upon re-run)

1. Real-time News Monitoring
    - Continuously checks for new industry news
    - Updates news context
    - Refreshes insights if significant changes
2. Forecast Updates
    - Recalculates forecasts with new data
    - Updates confidence intervals
    - Refreshes visualizations
3. Performance Monitoring
    - Tracks system performance
    - Monitors API usage
    - Logs user interactions
    - Handles errors gracefully

<u>This flow ensures that:</u>

- Each step builds upon previous steps
- All components work together seamlessly

- Users get comprehensive insights
- The system remains responsive
- Outputs are professional and actionable
- The experience is interactive and engaging

The agent maintains state throughout the session, allowing for:

- Quick filter updates
- Real-time visualization changes
- Seamless report generation
- Continuous news updates
- Efficient resource usage