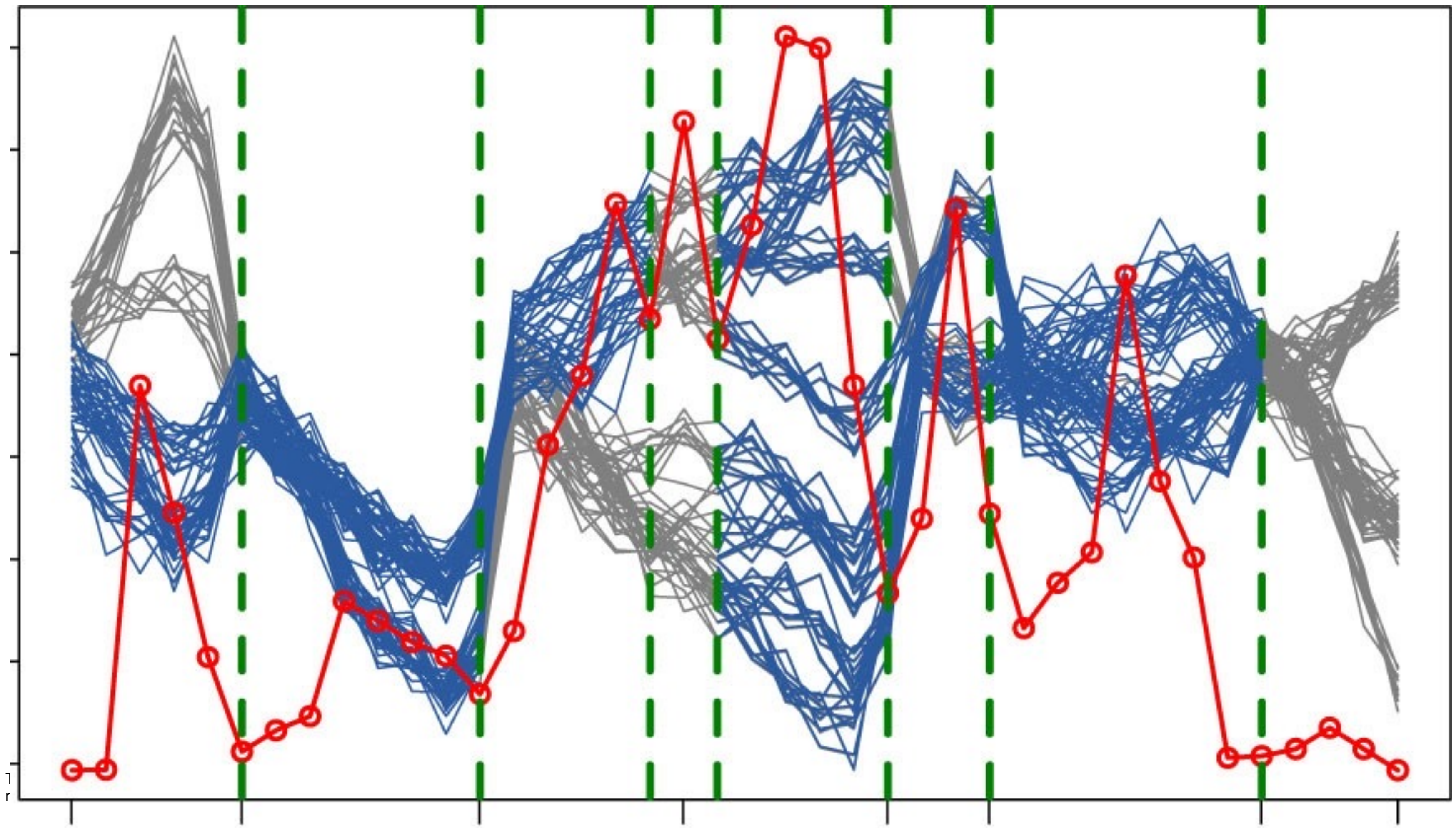
StreamStory: Where Time Series Meet Explainable AI – A Visual Journey Through Data

Jan Šturm, Maja Škrjanc, Tanja Zdolšek Draksler, Oleksandra Topal

Jožef Stefan Institute (JSI)

This Project has received funding from the European Union´s Research and Innovation programme under grant Agreement no 101092639

Introduction



Introduction

- StreamStory: A tool for interactive explorative analysis of multivariate time data
- Key challenges addressed:
  - Understanding complex temporal patterns
  - Handling large multivariate datasets
  - Delivering explainable insights

This project has received funding from the European Union’s Horizon research and innovation programme under Grant Agreement no 101092639

Why StreamStory?

Core Purpose:

- Makes complex data understandable for specialists in specific fields (domain experts)
- Acts as a bridge between two groups:
  - Data analysts (who process the data)
  - Domain experts (who need to use the insights)

Key Technical Features:

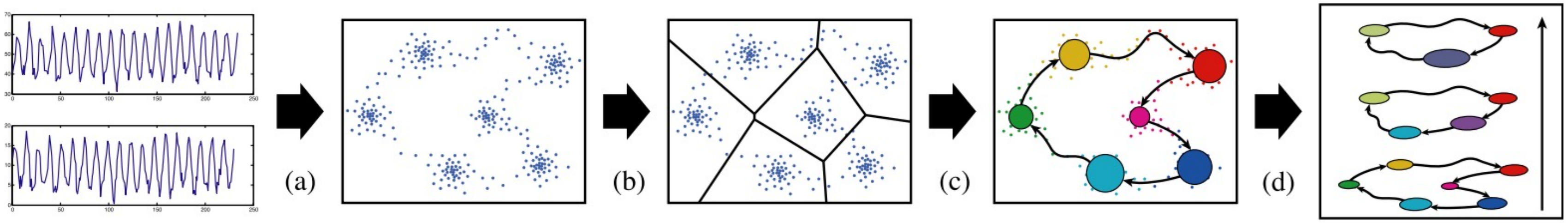
- Interactive visualization
- Multi-scale exploration
- AI-powered capabilities LLMs- automatically detects patterns in data

Main Benefits:

- Saves time in validation process
- Solves communication challenges
- Emphasizes trustworthiness and explainability of results
- Novel approach: Uses LLMs to interpret complex data patterns

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

How StreamStory works?



Multivariate input data streams

Convert time series (csv dataset) to point cloud

Construct states by partitioning space

Model transitions between states

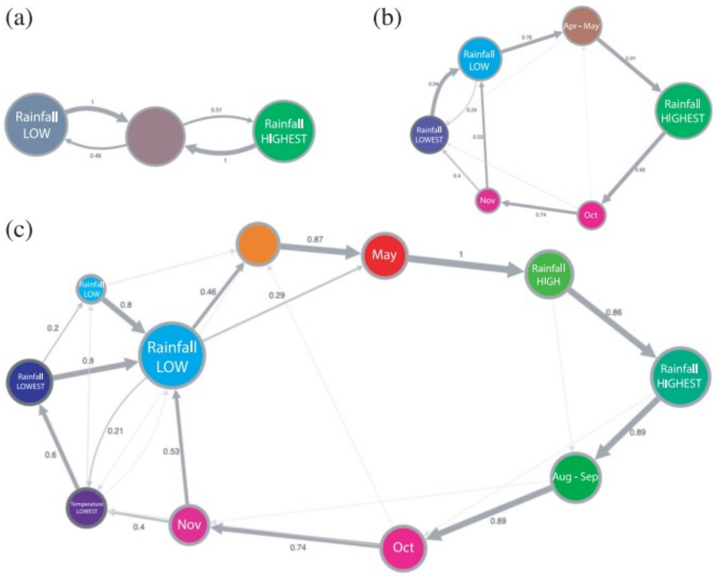
Create hierarchy of representations

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

User Interface

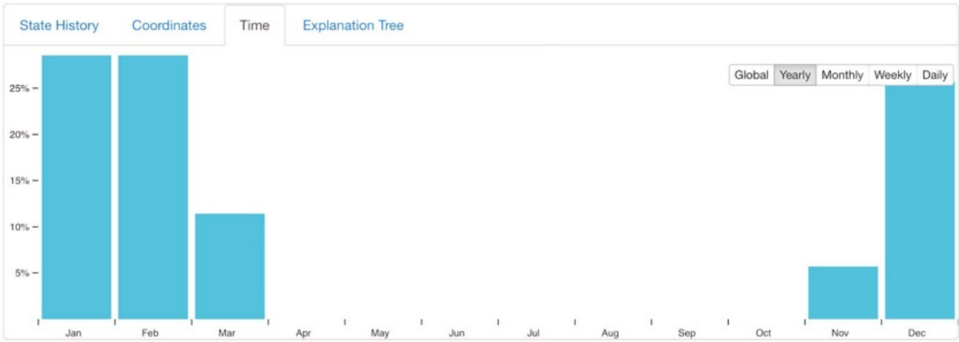
Visualizations

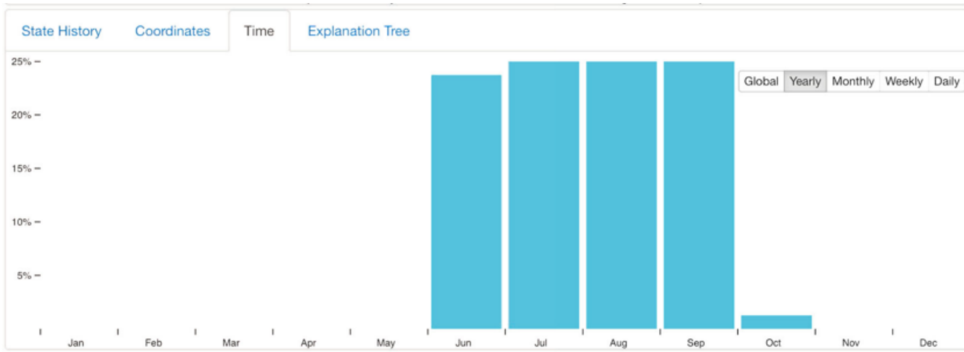
- Hierarchical Markov chain:
  - Attribute distributions
  - Domain expert selection
- State history
- Time distributions
- Explanation tree



Explainable AI Components

- Automatic state labeling
- Visual Analytics:
  - State color coding
  - Distributions
- LLM:
  - Currently in the testing phase
  - Extracting patterns from state history transitions
  - Using LLMs to explain these patterns in human-readable form
  - We will review an example of an LLM later

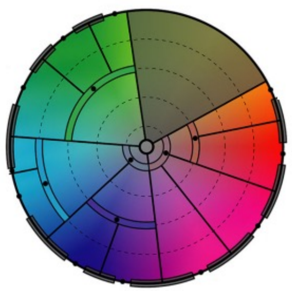


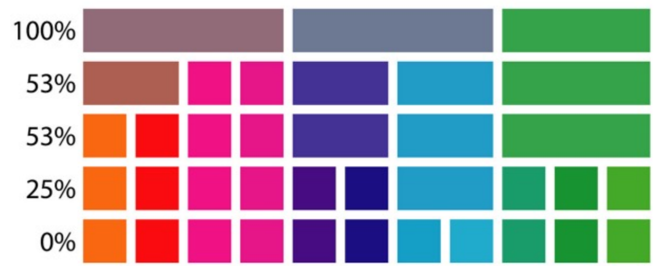


This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

State Color Coding

- Colors help visualize relationships between states across scales
- Works like a family tree:
  - Related states share similar colors
  - Distant states have distinct colors
- Color assignment process:
  - Initial colors distributed around color wheel
  - Each state gets range based on time spent in that state
  - When states merge at higher scales:
    - Colors blend together
    - Dominant states influence final color more
  - Color intensity indicates scale level:
    - Detailed states: saturated colors
    - Abstract states: more faded appearance
- Creates intuitive visual hierarchy for tracking state evolution





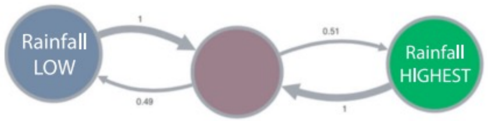
This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

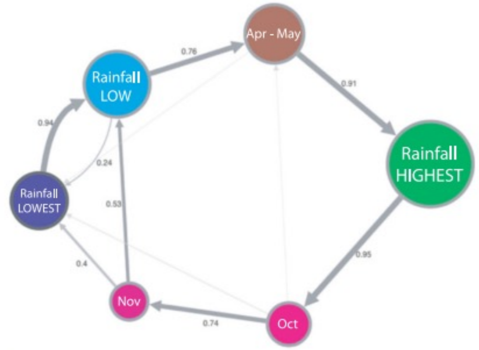
Case studies

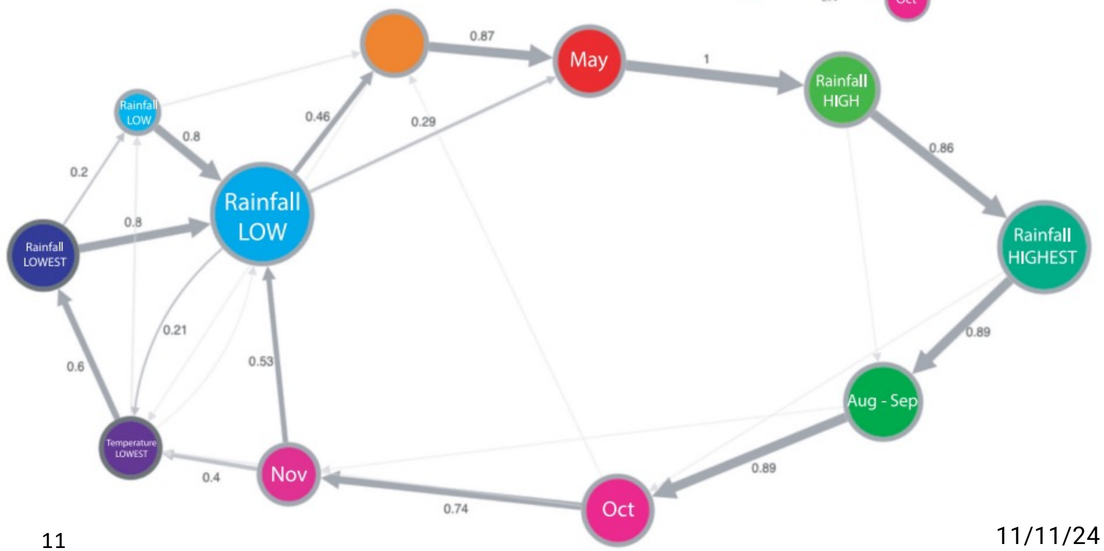
This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 1: Weather Data Analysis

- 20 years of UK weather data
- Variables: Temperature and rainfall monthly averages
- Discovered patterns:
  - Seasonal cycles
  - Weather state transitions
  - Temperature-rainfall relationships







This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

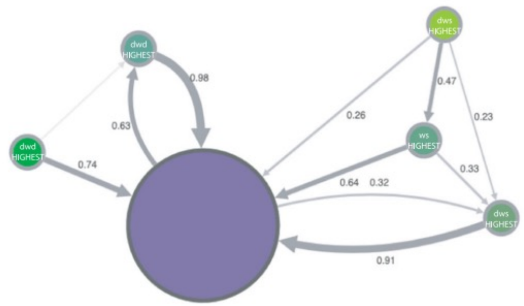
Case Study 2: GPS Tracking Analysis

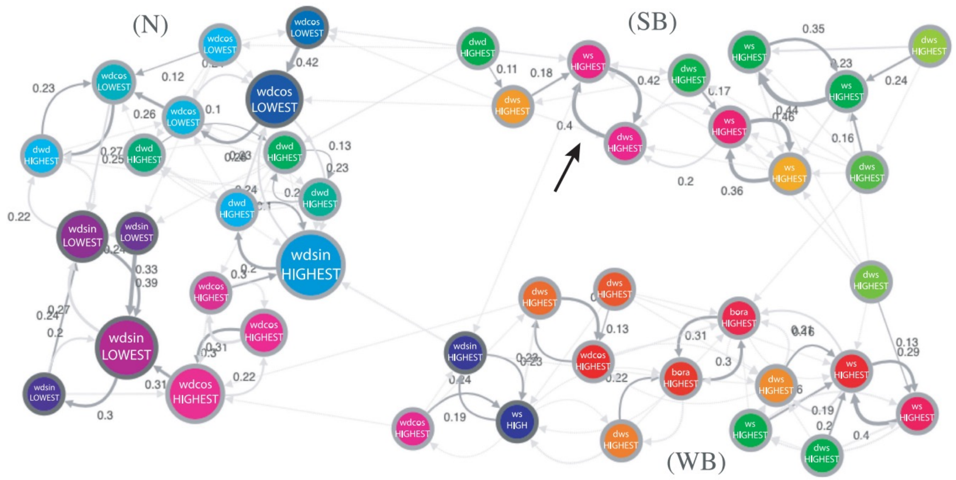
- Dataset: 3.5 years of GPS data (single person)
- Variables: GPS coordinates, timestamps
- Discovered patterns:
  - Home base centrality
  - Travel cycles
  - Geographic clusters

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 3: Wind

- Dataset:
  - Ajdovščina, Slovenia
  - Wind Bora measurements
  - March 2016 (10-minute intervals)
- Variables: Wind speed, direction, and their changes
- Main groups:
  - Calm winds (low speed, variable direction)
  - Strong Bora (SB) - perpendicular to ridge
  - Weak Bora (WB) - along valley direction



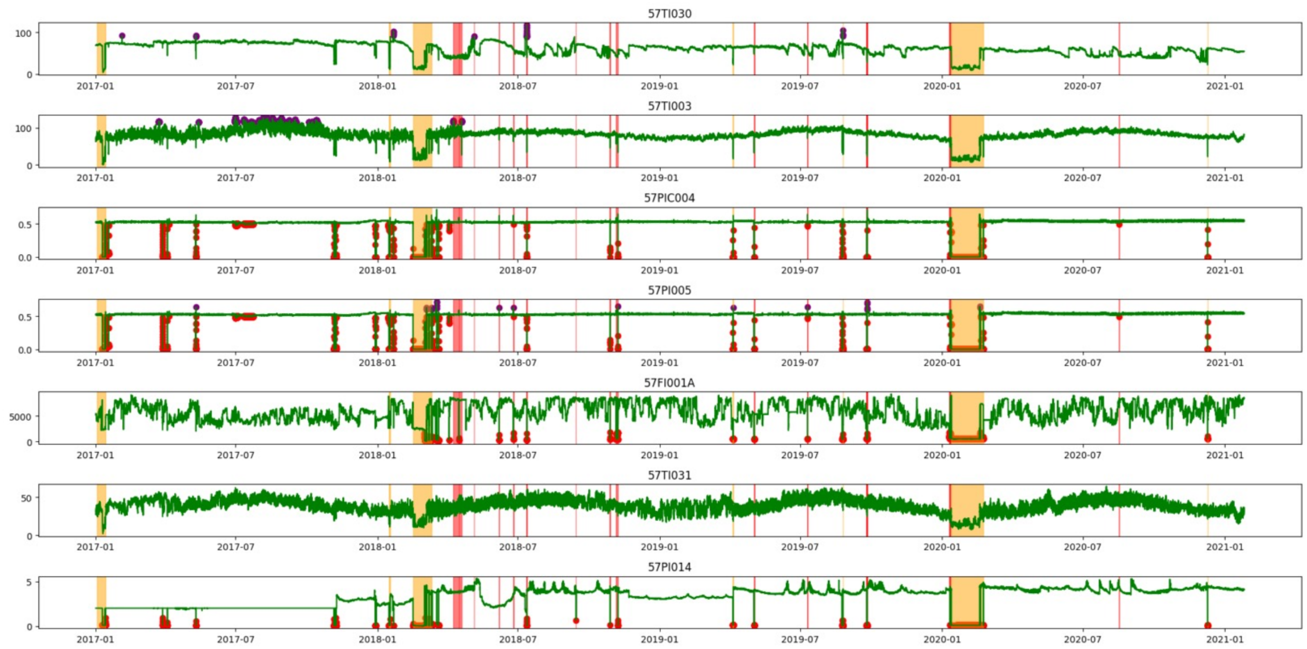


This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 4- FAME Pilot 7

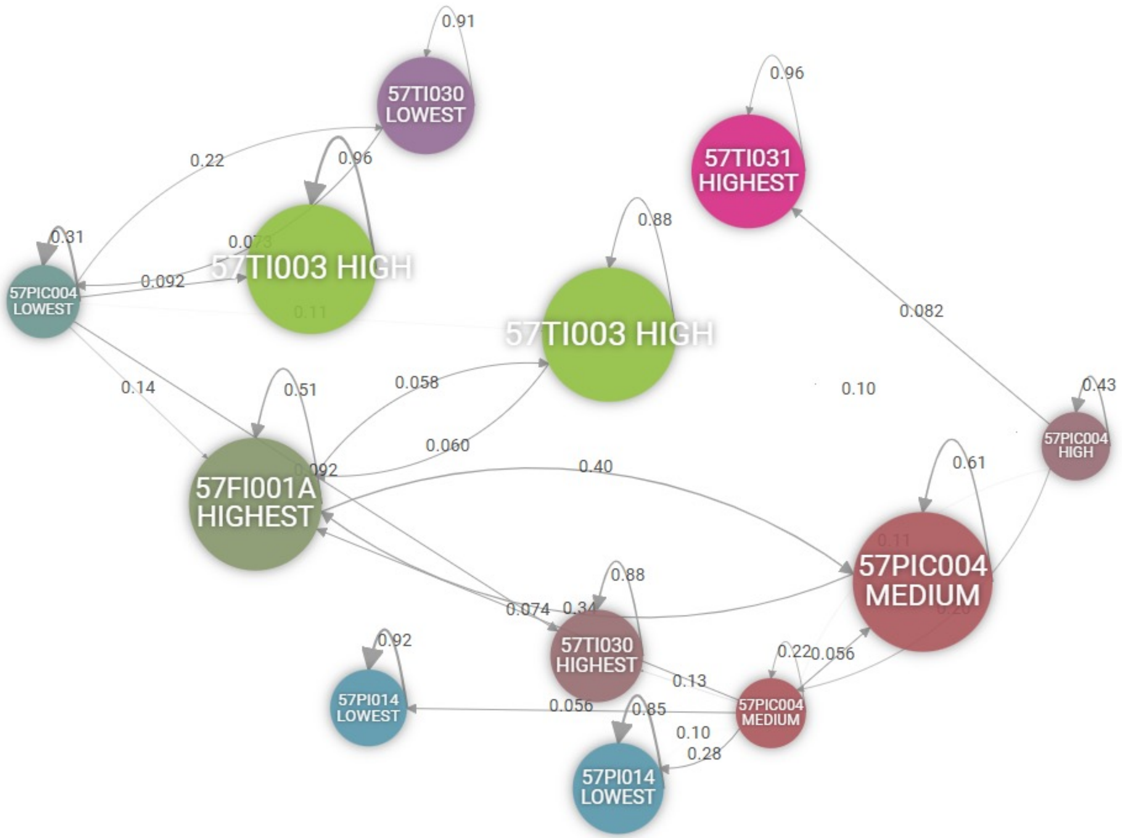
Dataset:

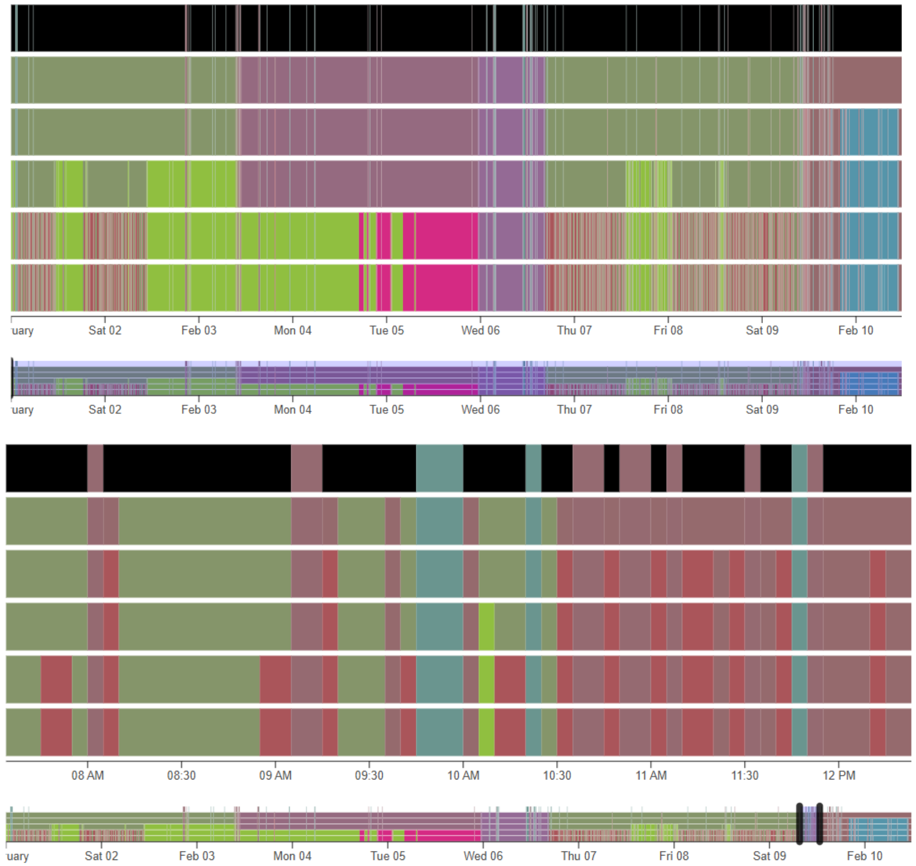
- Oil refinery data, including system failures and maintenance periods
- Sensors: temperature, pressure, flow for multiple machines



This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 4- FAME Pilot 7





This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 4- FAME Pilot 7

Current Challenges:

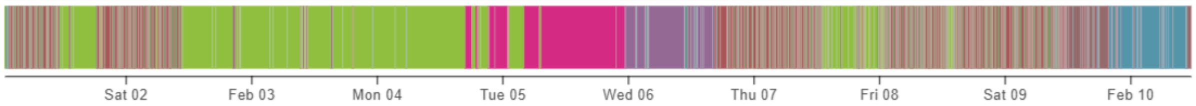
- Manual analysis of state transition patterns
- Reliance on expert knowledge

Proposed Solution:

- Leverage LLM to analyze Markov chains
- Input: Historical transition data and contextual descriptions
- Output: Detection of anomalous and recurring patterns and explanations of the patterns

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Case Study 4- FAME Pilot 7



Sequence of States	Type of Pattern	Time	Short Description	Intuitive Insight
57FI003 HIGH -> 57FI001A HIGHEST -> 57PI004 LOWEST	Recurrent	February 3-5, Morning	A sequence of high temperature followed by maximum pressure and lowest flow	Indicates a potential system stabilization phase, possibly due to a scheduled maintenance event or system reset.
57TI030 LOWEST -> 57PI014 LOWEST -> 57TI031 HIGHEST	Anomaly	February 6, Afternoon	A rapid shift from lowest temperature to highest temperature	Sudden temperature rise may indicate a malfunction or an unexpected external influence on the compressor system.
57PI004 LOWEST -> 57FI003 HIGH -> 57TI031 HIGHEST	Recurrent	February 7, 5-7 PM	Low flow followed by a high temperature increase	A common pattern during system ramp-up periods, reflecting the initial response to increased operational demand.
57TI031 HIGHEST -> 57PI014 LOWEST -> 57FI001A HIGHEST	Anomaly	February 8, Evening	Highest temperature followed by the lowest flow	May suggest a critical operational issue where the system fails to maintain pressure after peak temperature is reached, indicating a need for inspection.
57TI003 HIGH -> 57PI004 MEDIUM -> 57FI003 HIGH	Recurrent	February 9, 6-8 AM	Temperature and flow maintain higher levels with medium pressure	Reflects a standard operational cycle, likely during peak production hours, indicating system efficiency.
57FI001A HIGHEST -> 57PI014 LOWEST -> 57TI030 LOWEST	Anomaly	February 4, 11-1 PM	Transition from highest pressure to the lowest temperature	Indicates possible system instability, requiring attention to prevent potential failures during low-load conditions.

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Getting Started

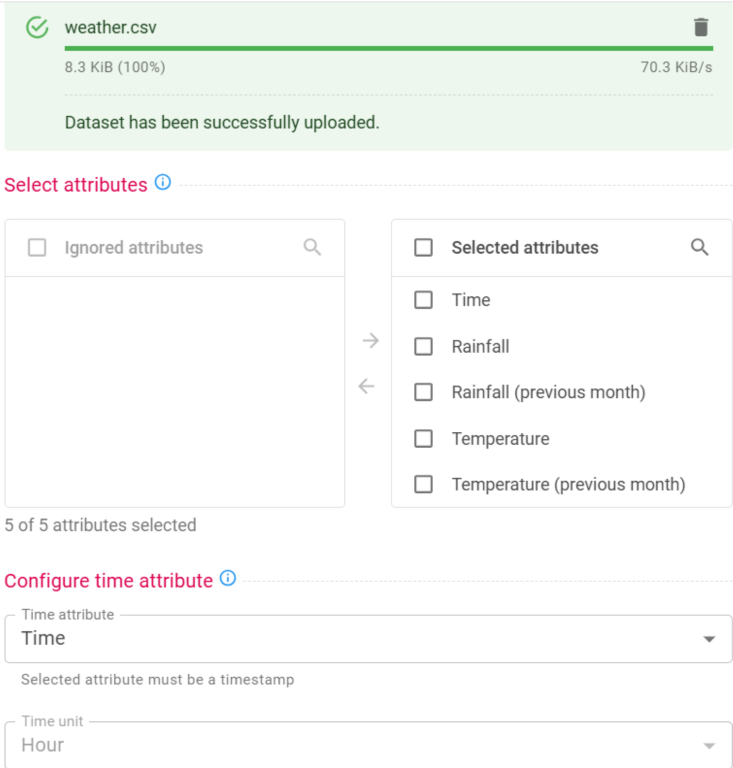
StreamStory integration

Getting Started

- Web application: http://streamstory.ijs.si
- Open source: https://github.com/E3-JSI/StreamStory2
- Data requirements:
  - CSV format
  - Aligned time series
  - Proper preprocessing:
    - Time series alignment
    - Missing value handling
    - Normalization
    - Feature engineering

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

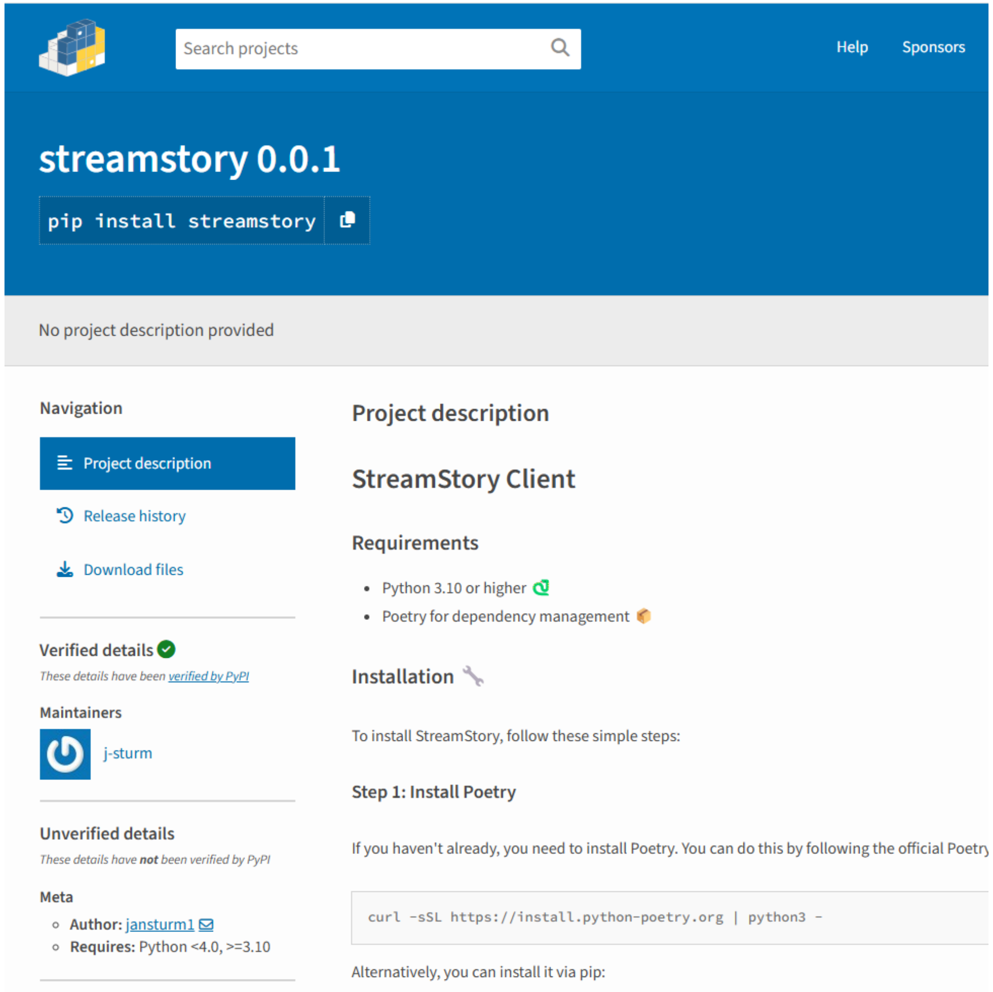
Build Models using UI



This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Build Models automatically

- API, swagger docs
- Python library



This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Live Demo

FEATURES

A qulitative multi-scale data analysis tool

StreamStory is a multi-scale data analysis tool for multivariate continuously time-varying data streams. It represents the data streams in a qualitative manner using states and transitions. Users can upload their own dataset or use one of the pre-loaded datasets. StreamStory can also be used as a monitoring tool, showing in real-time the state of the monitored process, activity and anomaly detection.

Exploratory data mining

A system for the analysis of multivariate time series. It computes and visualizes a hierarchical Markov chain model which captures the qualitative behavior of the systems' dynamics.

Multi-scale representation

The hierarchical model allows users to interactively find suitable scales for interpreting the data.

Real-time monitoring

Visualizes streaming data by mapping it to the hierarchical model. It can provide predictions and alarms for different behavior.

Free

Log in and get started. Check out our video presentation to see how its done and experiment with our example dataset.

Next Steps

- Validation of StreamStory models with domain experts
- Development and evaluation of various approaches and methods for using LLMs in XAI
- Applying these approaches in Pilot 7 of FAME project

This project has received funding from the European Union's Horizon research and innovation programme under Grant Agreement no 101092639

Q & A