

# Traffic Flow Prediction

ACODE - Tejas(61), Rajat(70), Tanay(94), Abhi(95)
CS-367 Artificial Intelligence
Indian Institute of Information Technology, Vadodara

# Overview

1. Problem statement

# Problem Statement



**Traffic Flow Prediction:** Urban traffic congestion increases travel times and fuel consumption. Current traffic management approaches are often inadequate for handling complex, dynamic traffic patterns. This project aims to develop a proactive, data-driven traffic management system using probabilistic models to predict traffic flows and congestion patterns.

# Research Papers




- Kim, J., & Wang, G. (2016). Diagnosis and Prediction of Traffic Congestion on Urban Road Networks Using Bayesian Network
- Sun, S., Zhang, C., & Yu, G. (2006). A Bayesian Network Approach to Traffic Flow Forecasting
- Horvat, R., Kos, G., & Ševrović, M. (2015). Traffic Flow Modelling on the Road Network in the Cities
- Adhitama, R. P., & Saputro, D. R. S. (2022). Hill climbing algorithm for Bayesian network structure

# Expected outcome



This project develops a comprehensive approach to traffic analysis and prediction using probabilistic models. The expected outcomes include identifying key congestion causes, generating short-term traffic flow predictions, and analyzing congestion patterns.

## 2 Approach

## Data Collection and Aggregation

Traffic parameters such as speed, intensity, travel time, and vehicle are utilized.

## Probabilistic Graphical Modeling

This model captures the probabilistic dependencies between factors influencing traffic, such as the time of day, vehicle size, and travel time.

## Traffic Forecasting

The model is used to predict traffic intensity at different intervals.

## Model Integration

Integrate the trained network into a real-time traffic management system.

## 3 Tech Stack

# Tech Stack



- **Language:** R used for statistical analysis and handling the probabilistic model.
- **Libraries/Frameworks:** bnlearn, readxl for applying probabilistic model and gathering data from databases.
- **Visualization:** Matplotlib for plotting and visualizing data, and prediction results.

4 Research Findings

# Findings



- Developed Bayesian Network (BN) model for traffic congestion analysis
- Data source: Historical traffic, incident, and weather data from Brisbane
- Model Components:
  - Input: Spatiotemporal factors (time, incidents, weather)
  - Output: Traffic metrics (flow, density, speed)
- Methodology:
  - Manual BN structure based on domain knowledge
  - Validation: Scoring functions and cross-validation
- Analysis techniques:
  - Odds ratios
  - Impact-probability matrices
- Outcome: Data-driven insights into congestion patterns and causes

# Use of Bayesian Network in Research



- Modeling probabilistic dependencies
  - BNs encode joint probability distributions
  - Variables represent factors affecting traffic conditions
- Probabilistic inference
  - BN models support both diagnostic and predictive reasoning
- Scenario analysis
  - Aids traffic managers in anticipating problematic conditions
- Capturing uncertainty
  - Provide probability distributions over predicted traffic states

# Methodology

- Data Collection and Preparation:
  - Historical traffic data from national traffic databases
  - Data aggregation to suitable time intervals
- Analysis and Inference
  - Calculate prior probabilities for various factors
  - Compute posterior probabilities to understand factor influences
  - Use Odds Ratio to identify leading causes of congestion

## 5 Functional Components

# Input



The input variables are based on historical traffic data and various factors related to traffic congestion. The input variables include:

- Traffic Speed
- Traffic Intensity
- Travel Time
- Vehicle Categories
- Day of the week
- Time of the day
- Street Network Data

# Output



The outputs of the research project, as presented, include:

- Short-term Traffic Predictions
- Identified Causes of Traffic Congestion
- Bayesian Network Model

# EndGoals



The end goals of the project are:

- Develop a method for predicting traffic congestion using a Bayesian Network model.
- Analyze traffic flow and make short-term predictions
- Create a data-driven framework for better understanding and addressing congestion.
- Improve traffic management by providing accurate and reliable predictions.

6 Algorithm

# Hill Climbing for Bayesian Networks



It is used to locate the BN shape by figuring out the conditional independence courting among variables.

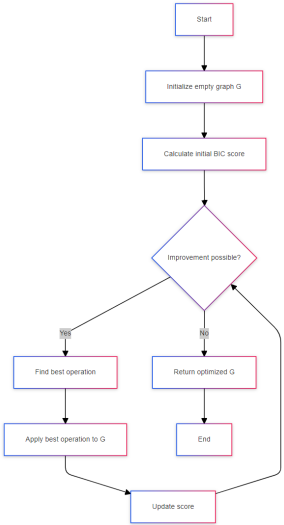
- **Input:** Dataset D of traffic flow variables
- **Output:** Optimized Bayesian Network structure Graph

# Hill Climbing Algorithm

1. Initialize
   - G = empty graph, score = BIC(G, D)
2. Repeat until no improvement or maximum iterations reached:
   - best_operation = null, best_score = score
   - For each possible operation (add, remove, or reverse an edge) in G:
     - G_new = Apply operation to G
     - If G_new is a valid DAG:
       new_score = BIC(G_new, D)
       If new_score > best_score:
       best_operation = operation, best_score = new_score
     - If best_operation is not null:
       G = Apply best_operation to G
       score = best_score
3. Return G as the optimized Bayesian Network structure:
   - LL = Calculate log-likelihood of G given D, k = Number of parameters in G , n = Number of samples in D
   - Return LL - (k*log(n)/2)

# Flowchart





7 Bayesian Network

# Data Preparation & Bayesian Network Structure Learning



1. Data Preprocessing:
   - Converted all variables to factors to ensure categorical treatment
2. Network Structure
   - Apply Hill-Climbing (HC) algorithm to learn a preliminary DAG
3. Scoring the Network
   - Evaluate dag using various scoring metrics to assess fit
4. Fitting the Bayesian Network
   - Fit the network to data to estimate conditional probability tables
5. Probabilistic Inference
   - Perform queries to compute conditional probabilities

# Bayesian Network