# Task 3: Traffic Volume Prediction

**User handout**

## How to run program:

1. **Database configuration –** 
   1. Download and install MySQL Server from <https://dev.mysql.com/downloads/mysql/>
   2. CREATE DATABASE traffic;
   3. CREATE TABLE `traffic`.`lanedata` (

`id` int(11) NOT NULL AUTO\_INCREMENT,

`lane\_id` int(11) NOT NULL,

`measurement\_date` datetime NOT NULL,

`speed` decimal(10,2) DEFAULT NULL,

`volume` int(11) DEFAULT NULL,

`occupancy` decimal(10,2) DEFAULT NULL,

`quality` int(11) DEFAULT NULL,

PRIMARY KEY (`id`)

) ENGINE=InnoDB AUTO\_INCREMENT=1190791 DEFAULT CHARSET=utf8;

1. Update database connection details in java program, config.properties “connectionUrl” field

Example –

connectionUrl =jdbc:mysql://localhost:3306/traffic?user=username&password=password123

1. Inserting data into database –
   1. Update **config.properties**:

Input file name in “**trainingdata**” field

Example – trainingData=lane\_data\_sample\_13\_11.csv



File location from “**dataDir**” field.

Example – dataDir=G:\\Research\\NIST\\data\\CATTLab\\lane

* 1. Run class **InsertIntoDatabase**.java (/TrafficVolumePrediction\_0.5/src/gov/nist/dataeval/data/InsertIntoDatabase.java)

1. Execute TrafficVolumePredictTest.java (/TrafficVolumePrediction\_0.5/test/gov/nist/dataeval/test/TrafficVolumePredictTest.java)

## Technical Details:

**Java Code Description –**

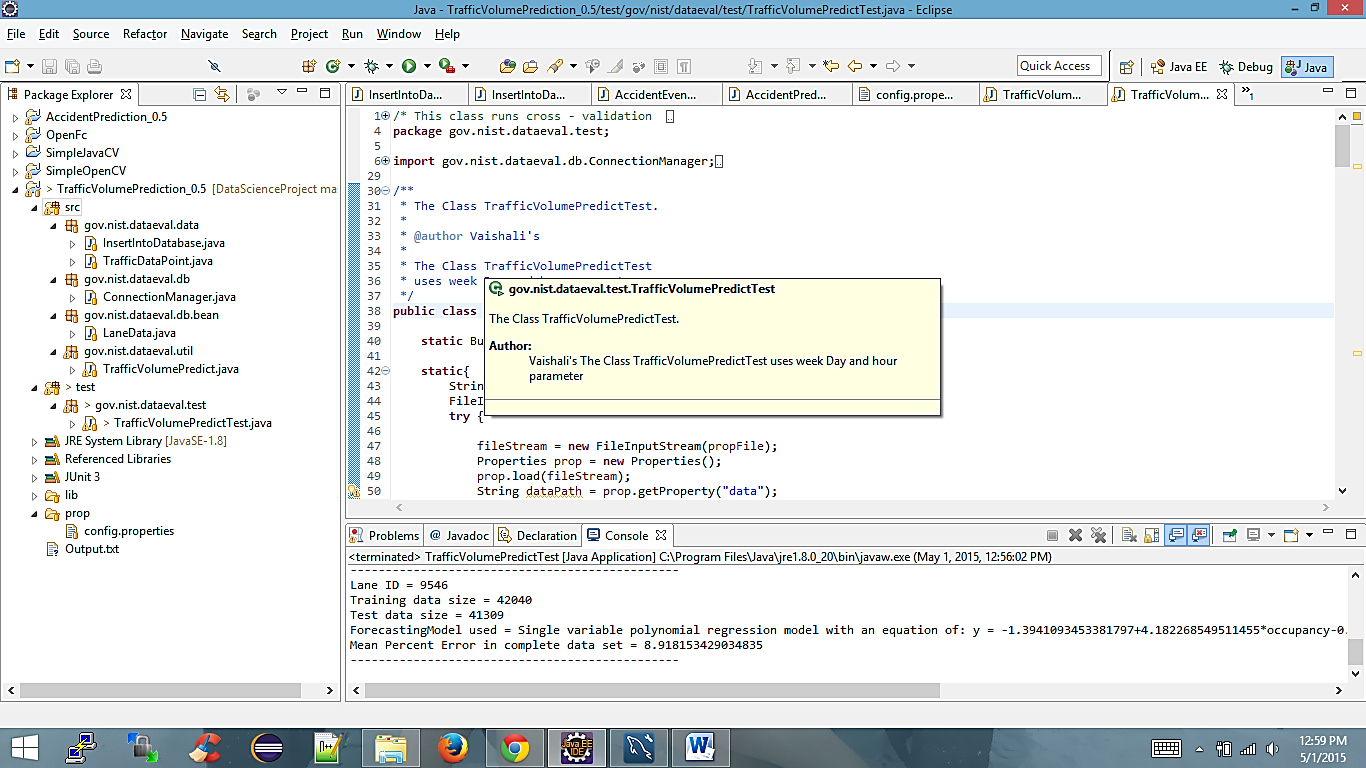


Fig. Code structure

**Config.properties –**

This file stores config parameters for project such as JDBC connection URLs, data location, training file name and test file name.

For inserting data into database –

**ConnectionManager.java –**

This is utility class to provide database connection using new ConnectionManager().getConnection() method.

**InsertIntoDataBase.java –**

This class has all the required JDBC code with preparedStatements to insert data from CSV file to MySQL Database.

String sql = "insert into traffic.lanedata(lane\_id,measurement\_date,speed,volume,occupancy,quality) values(?,?,?,?,?,?)";

**TrafficVolumePredict.java -**

This class uses OpenForecast open source library for machine learning to predict volume.

*learnAndPredictVolumeForLaneId*(laneId) this method takes lane Id as input, it fetches training data (e.g. Nov 2013 data) builds model using that training data (e.g. Nov 2013 data) and fetches testing data to test results.

While testing, we get predicted volume, we compare it against actual volume, and calculate Mean Percent Error using below formula –

Mean Percent Error (MPE) = \*100 %

## Results for training on Nov 2013 data and testing on Nov 2014 data:

Lane ID = 9873

Training data size = 41776

Test data size = 41336

ForecastingModel used = Single variable polynomial regression model with an equation of: y = 0.2209289722703055+1.1938838397680867\*occupancy+0.09636648597263575\*occupancy^2-0.008535737760676083\*occupancy^3

Mean Percent Error in complete data set = 7.368879427133733

-----------------------------------------------

Lane ID = 9544

Training data size = 40380

Test data size = 39770

ForecastingModel used = Single variable polynomial regression model with an equation of: y = -1.0390730373066546+3.4004047025847246\*occupancy+0.3309469880458526\*occupancy^2-0.07857147248169706\*occupancy^3+0.0052944763937361\*occupancy^4

Mean Percent Error in complete data set = 15.350766909730954

-----------------------------------------------

Lane ID = 9545

Training data size = 41198

Test data size = 40543

ForecastingModel used = Single variable polynomial regression model with an equation of: y = -2.130611301788455+4.789232747057021\*occupancy-0.07505801656220851\*occupancy^2-0.035914041758079045\*occupancy^3+0.002987357999749613\*occupancy^4

Mean Percent Error in complete data set = 14.6338455467035

-----------------------------------------------

Lane ID = 9546

Training data size = 42040

Test data size = 41309

ForecastingModel used = Single variable polynomial regression model with an equation of: y = -1.3941093453381797+4.182268549511455\*occupancy-0.04724187373349312\*occupancy^2-0.005130826001757049\*occupancy^3-0.0014277001639110636\*occupancy^4

Mean Percent Error in complete data set = 8.918153429034835

-----------------------------------------------