ĐẠI HỌC QUỐC GIA HÀ NỘI

ĐẠI HỌC QUỐC GIA HÀ NỘI

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



**NGUYỄN ĐỨC THẮNG**

**ÁP DỤNG MÔ HÌNH PHÂN LỚP VÀO DỰ ĐOÁN MẬT ĐỘ GIAO THÔNG**

**LUẬN VĂN THẠC SĨ HỆ THỐNG THÔNG TIN**

**Hà Nội - 2016**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

**NGUYỄN ĐỨC THẮNG**

**ÁP DỤNG MÔ HÌNH PHÂN LỚP VÀO DỰ ĐOÁN MẬT ĐỘ GIAO THÔNG**

Ngành: Công Nghệ Thông Tin

Chuyên ngành: Hệ thống Thông Tin

Mã số chuyên ngành: 60480104

**LUẬN VĂN THẠC SĨ HỆ THỐNG THÔNG TIN**

**NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS. TS. Nguyễn Trí Thành**

**Hà Nội – 2016**

**LỜI CAM ĐOAN**

Tác giả xin cam đoan kết quả đạt được trong luận văn là sản phẩm của riêng cá nhân Tác giả và được sự hướng dẫn khoa học của PGS. TS. Nguyễn Trí Thành, không sao chép lại của người khác. Trong toàn bộ nội dung của luận văn, những điều trình bày của cá nhân hoặc được tổng hợp của nhiều nguồn tài liệu. Tất cả các tài liệu tham khảo đều có xuất xứ rõ ràng và được trích dẫn hợp pháp.

Tác giả xin hoàn toàn chịu trách nhiệm và chịu mọi hình thức kỷ luật theo quy định cho lời cam đoan của mình.

|  |  |
| --- | --- |
|  | *Hà Nội, ngày tháng năm 2016*  **HỌC VIÊN**  **Nguyễn Đức Thắng** |

**LỜI CẢM ƠN**

Lời đầu tiên, em xin gửi lời cảm ơn chân thành và sâu sắc nhất tới PGS.TS Nguyễn Trí Thành, người thầy đã trực tiếp hướng dẫn tận tình và đóng góp những ý kiến quý báu cho em trong suốt quá trình thực hiện luận văn tốt nghiệp này.

Em xin gửi lời cảm ơn đến các thầy cô giáo trường Đại học Công nghệ - Đại học Công nghệ - Đại học Quốc gia Hà Nội, đã tận tâm truyền đạt những kiến thức quý báu làm nền tảng cho em trong công việc và cuộc sống.

Cuối cùng, em xin được cảm ơn cha mẹ, người thân, bạn bè và đồng nghiệp của em, những người đã luôn bên em, khuyến khích và động viên em trong cuộc sống và học tập.

|  |  |
| --- | --- |
|  | **HỌC VIÊN**  **Nguyễn Đức Thắng** |

MỤC LỤC

[DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT 7](#_Toc468060659)

[DANH MỤC BẢNG 8](#_Toc468060660)

[DANH MỤC HÌNH VẼ 9](#_Toc468060661)

[MỞ ĐẦU 10](#_Toc468060662)

[Chương 1: Giới thiệu chung 11](#_Toc468060663)

[1.1 Khái niệm Hệ thống giao thông thông minh 11](#_Toc468060664)

[1.2 Bài toán dự đoán mật độ giao thông 11](#_Toc468060665)

[1.3 Cơ sở phát triển và xây dựng bài toán Dự đoán mật độ giao thông 12](#_Toc468060666)

[Chương 2: Hệ thống giao thông thông minh. 14](#_Toc468060667)

[2.1 Mục tiêu của giải pháp Hệ thông giao thông thông minh (ITS) 14](#_Toc468060668)

[2.2 Mô tả giải pháp Hệ thống giao thông thông minh 17](#_Toc468060669)

[2.2.1 Hệ thống quản lý giao thông thông minh 19](#_Toc468060670)

[2.2.2 Hệ thống thông tin hành khách thông minh 20](#_Toc468060671)

[2.2.3 Hệ thống giao thông công cộng thông minh 21](#_Toc468060672)

[2.3 Lợi ích của Hệ thống giao thông thông minh 22](#_Toc468060673)

[2.4 Vai trò của Dự đoán mật độ giao thông trong Hệ thống giao thông thông minh 24](#_Toc468060674)

[Chương 3: Xây dựng mô hình dự đoán mật độ giao thông 26](#_Toc468060675)

[3.1 Bài toán phân lớp dữ liệu 26](#_Toc468060676)

[3.2 Mô hình dự đoán mật độ giao thông 27](#_Toc468060677)

[3.2.1 Mô hình đề xuất 27](#_Toc468060678)

[3.2.2 Sử dụng mô hình 27](#_Toc468060679)

[3.3 Các bước xây dựng mô hình Dự đoán mật độ giao thông 28](#_Toc468060680)

[3.3.1 Bước 1: Tạo dữ liệu thực nghiệm 28](#_Toc468060681)

[a. Lựa chọn đặc tính để sinh dữ liệu 28](#_Toc468060682)

[b. Thực hiện sinh dữ liệu 29](#_Toc468060683)

[3.3.2 Bước 2: Chia dữ liệu training và test 31](#_Toc468060684)

[3.3.3 Bước 3: Huấn luyện 33](#_Toc468060685)

[3.3.4 Bước 4: Thực hiện phân lớp 35](#_Toc468060686)

[3.4 Kết quả thực nghiệm 37](#_Toc468060687)

[3.4.1 Cài đặt môi trường thực nghiệm 37](#_Toc468060688)

[3.4.2 Thực nghiệm test 70-30 38](#_Toc468060689)

[a. Mô hình Decision Tree 38](#_Toc468060690)

[b. Mô hình Super Vector Machines (SVM) 38](#_Toc468060691)

[c. Mô hình Naive Bayes 39](#_Toc468060692)

[d. Mô hình Neural Network 39](#_Toc468060693)

[e. So sánh các kết quả chạy của SVM, Navies Bayes, J48 và Neural Network với số phân lớp là 6 39](#_Toc468060694)

[3.4.3 Thực nghiệm Test Cross validation 40](#_Toc468060695)

[a. Mô hình Decision Tree 40](#_Toc468060696)

[b. Mô hình Super Vector Machines (SVM) 41](#_Toc468060697)

[c. Mô hình Naive Bayes 41](#_Toc468060698)

[d. Mô hình Neural Network 42](#_Toc468060699)

[e. So sánh các kết quả chạy của SVM, Navies Bayes, J48 và Neural Network với số phân lớp là 6 42](#_Toc468060700)

[3.4.4 Thực nghiệm với các mức độ tắc đường khác nhau 43](#_Toc468060701)

[a. Thực nghiệp với số lớp (mức độ tắc đường) là 3 43](#_Toc468060702)

[b. Thực nghiệp với số lớp (mức độ tắc đường) là 4 43](#_Toc468060703)

[c. Thực nghiệp với số lớp (mức độ tắc đường) là 5 43](#_Toc468060704)

[d. Thực nghiệp với số lớp (mức độ tắc đường) là 6 43](#_Toc468060705)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TƯƠNG LAI 44](#_Toc468060706)

[TÀI LIỆU THAM KHẢO 45](#_Toc468060707)

# DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Từ viết tắt** | **Nghĩa đầy đủ** | **Ghi chú** |
| 1 | SVM | Support Vector Machine |  |
| 2 | ITS | Intelligent Transportation System |  |
| 3 | VMS | Variable Message Signs |  |
| 4 | CCTV | Closed Circuit Television |  |
| 5 | AVI | Automated Vehicle Identification |  |
| 6 | ITS | Intelligent Transport System |  |
| 7 | ISA | Intelligent Speed Adaptation |  |
| 8 | WIM | Weight In Motion |  |
| 9 | ETC | Electronic Toll Collection |  |
| 10 | CC | Cruise Control |  |
| 11 | PTP | Public Transport Priority |  |
| 12 | SCOOT | Split, Cycle and Offset Optimiser Technique |  |
| 13 | TDM | Travel Demand Management |  |
| 14 | HOV | High Occupancy Vehicle |  |
| 15 | WEKA | Waikato Environment for Knowledge Analysis |  |
| 15 | VSL | Variable Speed Limit |  |

# DANH MỤC BẢNG

[Bảng 2.1: Dự đoán mật độ giao thông với hệ thống giao thông thông minh 18](#_Toc469257446)

[Bảng 3.1: Cấu hình máy chủ trong thực nghiệm 40](#_Toc469257447)

[Bảng 3.2: Test 70-30 - Kết quả test mô hình Decision Tree 40](#_Toc469257448)

[Bảng 3.3: Test 70-30 - Kết quả test mô hình SVM 41](#_Toc469257449)

[Bảng 3.4: Test 70-30 - Kết quả test mô hình Naïve Bayes 41](#_Toc469257450)

[Bảng 3.5: Test 70-30 - Kết quả test mô hình Neural Network 42](#_Toc469257451)

[Bảng 3.6: Test 70-30 - So sánh kết quả phân lớp sử dụng SVM, Navies Bayes, J48 và Neural Network với bộ phân lớp là 6. 42](#_Toc469257452)

[Bảng 3.7: Test Cross Validation - Kết quả test mô hình Decision Tree 43](#_Toc469257453)

[Bảng 3.8: Test Cross Validation - Kết quả test mô hình SVM 44](#_Toc469257454)

[Bảng 3.9: Test Cross Validation - Kết quả test mô hình Naïve Bayes 44](#_Toc469257455)

[Bảng 3.10: Test Cross Validation - Kết quả test mô hình Neural Network 45](#_Toc469257456)

[Bảng 3.11: Test Cross - So sánh kết quả phân lớp sử dụng SVM, Navies Bayes, J48 và Neural Network. 45](#_Toc469257457)

[Bảng 3.12: Test Cross Validation - Kết quả test mô hình Decision Tree 46](#_Toc469257458)

[Bảng 3.13: Test Cross Validation - Kết quả test mô hình SVM 46](#_Toc469257459)

[Bảng 3.14: Test Cross Validation - Kết quả test mô hình Naïve Bayes 47](#_Toc469257460)

[Bảng 3.15: Test Cross Validation - Kết quả test mô hình Neural Network 47](#_Toc469257461)

# DANH MỤC HÌNH VẼ

[Hình 1:Mô hình dự đoán mật độ giao thông 27](#_Toc469257475)

[Hình 2: Test 70-30 – Dữ liệu training 35](#_Toc469257476)

[Hình 3: Test 70-30 – Dữ liệu Test 35](#_Toc469257477)

[Hình 4: Lựa chọn bộ dữ liệu huấn luyện 36](#_Toc469257478)

[Hình 5: Thông tin dữ liệu huấn luyện 36](#_Toc469257479)

[Hình 6: Thông tin tập Quan hệ hiện tại 36](#_Toc469257480)

[Hình 7: Thông tin các Attributes 37](#_Toc469257481)

[Hình 8: Thông tin chi tiết của các Attributes 37](#_Toc469257482)

[Hình 9: Màn hình phân lớp dữ liệu 38](#_Toc469257483)

[Hình 10: Chức năng chọn bộ phân lớp 38](#_Toc469257484)

[Hình 11: Các phương thức test 39](#_Toc469257485)

[Hình 12: Kết quả output phân lớp dữ liệu 40](#_Toc469257486)

[Hình 17: Test Cross Validation – Dữ liệu training 43](#_Toc469257487)

# MỞ ĐẦU

Ngày nay, với tốc độ đô thị hóa ngày càng trở thành một xu hướng phát triển cho hầu hết các thành phố trên thế giới nói chung và tại Việt Nam nói riêng, đã thúc đẩy nhu cầu đi lại và sử dụng các phương tiện giao thông của người dân ngày càng tăng cao. Điều này sẽ tạo ra các nhu cầu trong việc lựa chọn các tuyến giao thông thích hợp cho việc di chuyển giữa các địa điểm trong đô thị hoặc cả vùng ngoại ô. Với nhu cầu lựa chọn đó, việc hình thành các sản phẩm phần mềm hỗ trợ việc dự đoán mật độ tham gia giao thông tại một địa điểm xác định trong một khoảng thời gian xác định trước ngày càng trở nên thiết thực hơn bao giờ hết. Với mong muốn có thể đưa ra một hệ thống có thể dự đoán được mật độ giao thông dựa trên các thuật toán phân tích và xử lý dữ liệu, tác giả luận văn đã thực hiện việc nghiên cứu các thuật toán phân lớp một cách triệt để và tiến hành thực nghiệm hệ thống trên các dữ liệu đã thu thập được. Hệ thống sẽ dựa vào các thông tin và số liệu phân tích được sử dụng các mô hình phân lớp như SVM, Decision Tree, Bayer và Neural Network, sau đó trả về cho người dùng kết quả mật độ giao thông tại một vị trí xác định trong một khoảng thời gian xác định.

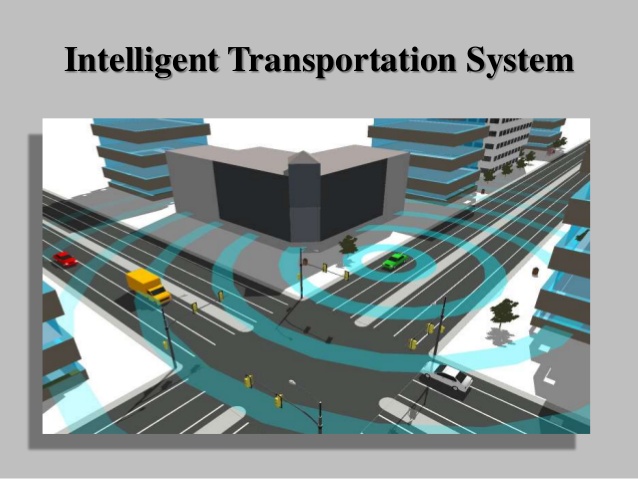
Trong luận văn nghiên cứu này, tác giả trình bày trong 3 chương với nội dung được tóm tắt như sau:

* Chương 1. **Giới thiệu chung** trình bày các khái niệm về thống giao thông thông minh và đưa ra bài toán về dự đoán mật độ giao thông. Cơ sở phát triển và xây dựng bài toán cũng được trình bày trong phần cuối của chương này.
* Chương 2. **Hệ thống giao thông thông minh** nghiên cứu chi tiết các phần như mục tiêu, mô tả giải pháp, lợi ích cũng như vai trò chính của Hệ thống giao thông thông minh.
* Chương 3. **Xây dựng mô hình dự đoán mật độ giao thông** đưa ra các nghiên cứu về bài toán phân lớp dữ liệu và mô hình dự đoán mật độ giao thông. Trong chương này, tác giả cũng đưa ra các bước để có thể xây dựng mô hình dự đoán mật độ giao thông và các kết quả thực nghiệm được tiến hành bởi chính tác giả.
* Phần **Kết luận và hướng phát triển tương lai** trình bày những kết quả đã đạt được và hạn chế trong luận văn. Các vấn đề còn hạn chế sẽ được giải quyết trong hướng phát triển tương lai của luận văn.

# Chương 1: Giới thiệu chung

## Khái niệm Hệ thống giao thông thông minh

Hệ thống Giao thông Thông minh (lntelligent Transport System - ITS) là hệ thống điều khiển giao thông thông qua các thiết bị điện tử, các hệ thống giám sát với mục địch giảm thiểu tối đa tai nạn, sự cố giao thông và tận dụng được tối đa điều kiện cơ sở vật chất.



Các biện pháp của Hệ thống giao thông thông minh - ITS được chia thành 3 nhóm:

* **Hệ thống quản lý giao thông thông minh:** chức năng chính của hệ thống này là quản lý và phân tích thông tin lưu lượng giao thông và đưa ra các biện pháp để giảm thiểu các vấn đề có thể xay ra. Hệ thống này bao gồm việc như kiểm soát tín hiệu, đường cao tốc và hệ thống quản lý lưu lượng giao thông, cấp phép điện tử, các hệ thống quản lý sự cố, thu phí điện tử và giá cả, hệ thống thực thi giao thông và thích ứng tốc độ thông minh.
* **Hệ thống quản lý hành khách thông minh:** hệ thống này giúp nâng cao kiến ​​thức cho hành khách trong việc tham gia giao thông và bao gồm các hệ thống thông tin hành khách, hệ thống hướng dẫn đường được tích hợp trực tiếp trong xe ô tô, hệ thống hướng dẫn đỗ xe vào các vị trí còn trống, cung cấp các cơ sở dữ liệu bản đồ kỹ thuật số và hệ thống tin nhắn về trạng thái giao thông.
* **Hệ thống giao thông công cộng thông minh:** bao gồm các biện pháp giao thông thông minh nhằm cải thiện hiệu suất vận tải công cộng. Hệ thống này bao gồm các loại xe thông minh, tốc độ của xe thích ứng một cách thông minh, hệ thống quản lý đội xe vận chuyển, hệ thống thông tin hành khách quá cảnh, hệ thống thanh toán điện tử, cấp phép điện tử, các hệ thống quản lý nhu cầu giao thông vận tải và ưu tiên giao thông công cộng.

## Mục tiêu của giải pháp Hệ thông giao thông thông minh (ITS)

Mục tiêu quan trọng: ITS có thể thuận lợi cung cấp một loạt các giải pháp, hướng phát triển, nhằm mang lại các điều kiện thuận lợi về giao thông cho những người di chuyển hoặc những người sống trong khu vực mà hệ thống quản lý . Có sáu mục tiêu chính / lợi ích đã được xác định trong các tài liệu quốc:

**An toàn:** Một mục tiêu rõ ràng của hệ thống giao thông là cung cấp một môi trường an toàn cho việc đi lại trong khi tiếp tục phấn đấu để cải thiện hiệu suất của hệ thống. Mặc dù không mong muốn, tai nạn và tử vong rất hay xảy ra không thể tránh khỏi. Một số dịch vụ ITS nhằm giảm thiểu nguy cơ tai nạn xảy ra. Mục tiêu này tập trung vào việc giảm số tai nạn và giảm xác suất của một tử vong nên một vụ tai nạn xảy ra. biện pháp điển hình về hiệu quả sử dụng để xác định số lượng hoạt động an toàn bao gồm tỷ lệ tai nạn tổng thể, tỷ lệ tai nạn tử vong và tỷ lệ tai nạn chấn thương. dịch vụ ITS cũng cần phải phấn đấu giảm tỷ lệ tai nạn của một cơ sở hoặc hệ thống. Tỷ lệ tai nạn thường được tính toán về tai nạn mỗi năm, được tính dựa trên mỗi 1 triệu km hoặc trên phạm vi 10,000 dân.

**Di động:** Cải thiện tính di động (và độ tin cậy) bằng cách giảm sự chậm trễ và thời gian đi lại là một mục tiêu chính của nhiều thành phần ITS. Sự chậm trễ có thể được đo bằng nhiều cách khác nhau, tùy thuộc vào loại hệ thống giao thông được phân tích. Sự chậm trễ của một hệ thống thường được tính bằng giây hoặc vài phút của sự chậm trễ mỗi xe. Ngoài ra, sự chậm trễ cho người sử dụng của hệ thống có thể được đo bằng số người-giờ. Sự chậm trễ về vận chuyển hàng có thể được đo lường trong thời gian qua thời gian đến dự kiến ​​của lô hàng. Sự chậm trễ này cũng có thể được đo bằng cách quan sát số điểm dừng kinh nghiệm của trình điều khiển trước và sau khi dự án được triển khai hoặc triển khai.

Du lịch thời gian biến thiên cho thấy sự biến đổi trong thời gian du lịch tổng thể từ một nguồn gốc đến một điểm đến trong hệ thống, bao gồm bất kỳ chuyển phương thức hoặc trên đường đi dừng. Biện pháp này hiệu quả có thể dễ dàng được áp dụng cho vận tải đa phương (hàng hoá) phong trào cũng như du lịch cá nhân. Giảm sự thay đổi của thời gian đi cải thiện độ tin cậy của thời gian đến ước tính rằng khách du lịch hoặc công ty sử dụng để đưa ra quyết định lập kế hoạch và lịch trình. Bằng cách cải thiện hoạt động và sự cố phản ứng, và cung cấp thông tin về sự chậm trễ, dịch vụ ITS có thể làm giảm sự thay đổi của thời gian đi lại trong mạng lưới giao thông. Ví dụ, sản phẩm thông tin du lịch có thể được sử dụng trong việc lập kế hoạch chuyến đi để giúp lái xe thương mại lại tuyến đường xung quanh khu vực tắc nghẽn dẫn đến ít biến đổi trong thời gian đi lại.

**Hiệu quả:** Nhiều thành phần ITS tìm kiếm để tối ưu hóa hiệu quả của các cơ sở hiện có và sử dụng các “hướng xử lý đúng” để tính di động và thương mại điện nhu cầu có thể được đáp ứng trong khi giảm sự cần thiết phải xây dựng hoặc mở rộng cơ sở. Điều này được thực hiện bằng cách tăng năng lực hiệu quả của hệ thống giao thông vận tải. năng lực hiệu quả là "tỷ lệ tối đa tiềm năng mà người hoặc phương tiện có thể đi qua một liên kết, nút hoặc mạng dưới một hỗn hợp đại diện của điều kiện đường bộ", bao gồm "thời tiết, sự cố và sự thay đổi trong mô hình nhu cầu giao thông". Năng lực, theo định nghĩa từ các nghiên cứu, năng lực là "lượng tối đa mà người hoặc xe hợp lý có thể được dự kiến ​​sẽ đi qua một điểm cho trước của một làn đường hoặc đường trong một khoảng thời gian nhất dưới hiện hành điều kiện đường bộ, giao thông và kiểm soát”. Sự khác biệt lớn giữa năng lực và năng lực hiệu quả là công suất thường được đo trong điều kiện điển hình cho cơ sở, chẳng hạn như thời tiết và mặt đường điều kiện tốt, không có sự cố ảnh hưởng đến hệ thống, trong khi năng lực hiệu quả có thể thay đổi tùy theo các điều kiện và việc sử dụng quản lý và chiến lược hoạt động.

Thông lượng được định nghĩa là số lượng người, hàng hoá, phương tiện đi qua một phần đường hoặc mạng cho mỗi đơn vị thời gian. Tăng thông lượng đôi khi ngộ của tăng năng lực hiệu quả. Dưới những điều kiện nhất định, nó có thể phản ánh số lượng tối đa của du khách có thể được cung cấp bởi một hệ thống giao thông. Thông được dễ dàng hơn đo so với năng lực và hiệu quả, do đó, có thể được sử dụng như một biện pháp thay thế khi phân tích hiệu suất của một dự án ITS. Người đọc cần lưu ý rằng hoàn cảnh địa phương ảnh hưởng đến năng lực địa phương, cũng như thông lượng đo.

**Năng suất:** ITS thường làm giảm chi phí vận hành và cho phép cải thiện năng suất. Ngoài ra, lựa chọn thay thế ITS có thể thấp hơn mua lại và vòng đời của chi phí so với các kỹ thuật cải thiện giao thông truyền thống. Biện pháp hiệu quả cho mục tiêu này là tiết kiệm chi phí như là kết quả của việc thực hiện ITS. Một cách khác để xem những tiết kiệm chi phí là để định lượng các khoản tiết kiệm chi phí giữa các giải pháp truyền thống và ITS để giải quyết vấn đề.

**Năng lượng và môi trường:** chất lượng không khí và năng lượng tác động của dịch vụ ITS rất quan trọng, đặc biệt đối với các khu vực không đạt chất lượng. Trong hầu hết các trường hợp, lợi ích môi trường chỉ có thể được ước tính bằng cách sử dụng phân tích và mô phỏng. Các vấn đề liên quan đến đo lường khu vực bao gồm các tác động nhỏ của từng dự án và số lượng lớn các biến ngoại sinh bao gồm thời tiết, đóng góp từ các nguồn không di động hoặc các khu vực khác, và bản chất thời gian phát triển của ô nhiễm ôzôn. nghiên cứu quy mô nhỏ nói chung cho thấy tác động tích cực đến môi trường, và những tác động này là kết quả của dòng chảy mượt mà và hiệu quả hơn trong các hệ thống giao thông. Tuy nhiên, tác động môi trường của du khách phản ứng để triển khai quy mô lớn trong dài hạn chưa được hiểu rõ. Việc giảm mức khí thải và tiêu thụ năng lượng đã được xác định là biện pháp hiệu quả cho mục tiêu này.

**Sự hài lòng của khách hàng:** Do nhiều dự án và chương trình ITS đã được phát triển đặc biệt để phục vụ công chúng, điều quan trọng để đảm bảo người dùng (khách hàng) mong đợi được đáp ứng hoặc vượt qua. Biện pháp làm hài lòng khách hàng và đặc trưng cho khoảng cách giữa kỳ vọng và kinh nghiệm của người sử dụng liên quan đến dịch vụ hoặc sản phẩm. Các câu hỏi trung tâm trong một đánh giá sự hài lòng của khách hàng là: "Liệu các sản phẩm cung cấp đầy đủ giá trị (hoặc lợi ích) để đổi lấy sự đầu tư của khách hàng, dù đầu tư được tính bằng tiền hoặc thời gian?" Kết quả tiêu biểu báo cáo trong việc đánh giá tác động của sự hài lòng của khách hàng với một sản phẩm hoặc dịch vụ bao gồm nhận biết sản phẩm, kỳ vọng của lợi ích sản phẩm (s), sử dụng sản phẩm, đáp ứng (ra quyết định hoặc thay đổi hành vi), thực hiện các lợi ích và đánh giá giá trị. Mặc dù sự hài lòng là khó đo lường trực tiếp, các biện pháp liên quan đến sự hài lòng có thể được quan sát bao gồm cả số lượng du lịch trong các chế độ khác nhau, và chất lượng dịch vụ, cũng như số lượng khiếu nại và / hoặc khen nhận được bởi các nhà cung cấp dịch vụ.

 Ngoài sử dụng hoặc sự hài lòng của khách hàng, ITS là cần thiết để đánh giá sự hài lòng của các nhà cung cấp hệ thống giao thông hoặc người quản lý. Ví dụ, nhiều dự án ITS được thực hiện để cải thiện sự phối hợp giữa các bên liên quan khác nhau trong lĩnh vực giao thông vận tải. Trong các dự án như vậy, điều quan trọng là để đo lường sự hài lòng của các nhà cung cấp vận chuyển để đảm bảo việc sử dụng tốt nhất nguồn vốn hạn chế. Một cách để đo lường hiệu quả của một dự án như vậy là để khảo sát các nhà cung cấp vận chuyển trước và sau khi dự án đã được thực hiện để xem nếu phối hợp được cải thiện. Nó cũng có thể có thể để mang lại cùng các nhà cung cấp từ mỗi nhóm liên quan để đánh giá sự hài lòng của họ với hệ thống trước và sau khi thực hiện một dự án ITS.

## Bài toán dự đoán mật độ giao thông

Với tình hình phát triển về kinh tế hiện tại, số lương phương tiện giao thông đặc biệt là ô tô, xe máy đang ngày phát triển với tốc độ lớn về số lượng. Tuy nhiên đi cùng với tốc độ phát triển của phương tiên giao thông thì hạ tầng giao thông lại chưa phát triển một cách tương xứng. Trong các khung giờ và khung đường nhất định, hiện tượng ùn ứ thường xuyên xảy ra, gây ra khó chịu cho người tham gia giao thông đồng thời cũng gây nên các tác động không tốt với môi trường. Người tham gia giao thông trước khi di chuyển không thể biết được đoạn đường nào sẽ tắc và đoạn đường nào không vào khoảng thời gian đó nên họ không biết nên lựa chọn khung đường nào là thích hợp. Hiện tại, ngoài các đèn tín hiệu để phân bổ giao thông, các chiến sĩ công an cũng được sắp xếp xuất hiện và hướng dẫn giao thông ở các địa điểm thường xuyên xảy ra tắc. Tuy nhiên với giới hạn về người và chi phí, việc bố trí người ở các điểm xảy ra tắc chưa được hoàn toàn tối ưu. Nhiều điểm tắc mạnh nhưng ít công an giao thông và ngược lại nhiều điểm ít tắc lại được bố trí nhiều công an giao thông hơn.

Với tình hình giao thông như trên, nếu có thể dự đoán được mật độ giao thông tại các khung đường và khung giờ nhất định sẽ góp phần giảm thiểu được ùn tắc và tận dụng tối đa nguồn lực. Người tham gia giao thông có thể dựa vào mật độ giao thông được dự đoán để lựa chọn khung đường di chuyển thích hợp, tránh những nơi sẽ đông và di chuyển theo lộ trình ít phương tiện hơn. Nếu dự đoán trước được mật động giao thông, các chiến sĩ công an cũng sẽ được phân bố vị trí một cách hiệu quả hơn. Những điểm được dự đoán là sẽ tắc mạnh sẽ được dồn nhiều nguồn lực hơn những điểm khác.

Hệ thống giao thông thông minh đang là mục tiêu phát triển của thế giới và việc dự đoán được mật độ giao thông là một phần rất quan trọng trong hệ thống này. Hệ thống giao thông thông minh sẽ dựa vào thông tin được dự đoán để thông báo và đưa ra các khuyến cáo cho người tham gia giao thông nhằm tận dụng được tối đa cơ sở hạ tầng và giảm thời gian bị ảnh hưởng bởi việc tắc đường và ùn ứ xuống mức thấp nhất.

## Cơ sở phát triển và xây dựng bài toán Dự đoán mật độ giao thông

Theo thống kê của sở giao thông, tại Hà Nội, 3 tháng đầu năm 2016 có tất cả 34 điểm ùn tắc giao thông thường xuyên, trong đó có:

* 7 điểm xuất hiện ở các khu nhà ở cao tầng, nơi có mật độ dân cư cao
* 11 điểm xuất hiện ở các công trình đang xây dựng
* 5 điểm thường xuyên ùn tắc do là tuyến đường trọng điểm, trục đường chính với lưu lượng giao thông lớn

Theo các thống kê từ VOV giao thông thì việc tắc đường thưởng xảy ra theo khung giờ nhất định:

* 6h30-8h00: đây là khung giờ đi làm của người lớn và đi học của học sinh sinh viên nên khung giờ này có độ tắc rất cao.
* 11h00-12h00: đây là khung giờ nghỉ trưa, khung giờ thường xuyên xảy ra hiện tượng ùn ứ do lượng xe cộ trong khung giờ này là rất cao.
* 16h30-18h00: đây là giờ tan học của học sinh, sinh viên và tan làm của người lớn nên khung giờ này thường xuất hiện tắc nghẽn nghiêm trọng ở rất nhiều khung đường khác nhau.

Như vậy, dựa vào các số liệu thống kê ta có thể thấy việc ùn ứ, tắc đường này thường có tính quy luật, ví dụ như tại các thời điểm bắt đầu đi làm buổi sáng hoặc tan tầm là các thời điểm thường xuyên xảy ra tắc đường. Và tại các nút giao thông quan trọng, tình trạng tắc đường cũng thường xuyên xảy ra.

Chúng ta có thể tập hợp các thông thống kê này lại và xây dựng nên một bộ dữ liệu tương ứng. Bộ dữ liệu này được xây dựng dựa vào các đặc điểm, các khoảng thời gian tắc đường, địa điểm hay xảy ra tắc đường và hệ thống sẽ sử dụng bộ dữ liệu này để phân tích và đưa ra các dự báo một cách tương đối chính xác cho những người tham gia giao thông.

# Chương 2: Hệ thống giao thông thông minh.

## Mô tả giải pháp Hệ thống giao thông thông minh

Hệ thống giao thông thông minh (ITS) là một lĩnh vực rất rộng lớn. ITS có thể được sử dụng trong một số lĩnh vực như điều khiển đèn giao thông để quản lý sự cố, quản lý thông tin hành khách và hỗ trợ điều khiển để thực thi giới hạn tốc độ thông minh. Để có thể tạo ra một cấu trúc hoàn chỉnh cho một lĩnh vực rộng lớn như Hệ thống giao thông thông minh là thực sự khó khăn vì mỗi cấu trúc trong một lĩnh vực riêng biệt đều là một thách thức lớn đối với các nhà chức trách.

An toàn, nhanh nhẹn, hiệu quả, năng suất, năng lượng và môi trường, và sự hài lòng của khách hàng là những thế mạnh của hệ thống ITS.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Hệ thống quản lý giao thông thông minh | Hệ thống quản lý khách hàng thông minh | Hệ thống phương tiện giao thông công cộng thông minh |
| An toàn | * Giới hạn tốc độ   Quản lý làn đường   * Quản lý tai nạn * Hệ thống cảnh báo   Camera giám sát   * Xe tự động * Kiểm soát tốc độ thông minh   Quản lý trọng tải | Hệ thống định vị   * Hướng dẫn đỗ xe * Kiểm soát hành trình   Hệ thống hộp đen  Xe tự động  Hệ thống bến đỗ   * Cảnh báo khoảng cách | * Quản lý theo đội   Hệ thống định vị  Vé điện tử  Camera giám sát   * Phương tiện vận chuyển cao tốc * Xe tự động * Kiểm soát tốc độ thông minh * Cảnh báo khoảng cách |
| Di động và hiệu quả | * Giới hạn tốc độ   Quản lý làn đường   * Quản lý tai nạn * Hệ thống cảnh báo   Camera giám sát  Đo khoảng cách đường nối   * Điều khiển phương tiện   Thu phí điện tử   * Thông tin thồi gian thực * Hướng dẫn đỗ xe | Hệ thống định vị   * Hướng dẫn đỗ xe * Kiểm soát hành trình * Hệ thống cảnh báo | Xe ưu tiên   * Quản lý theo đội   Hệ thống định vị  Vé điện tử  Hệ thống tương tác   * Phương tiện vận chuyển cao tốc * Thông tin thời gian thực |
| Sự thảo mãn của khách hàng | Camera giám sát  Quản lý làn đường   * Hệ thống cảnh báo   Thu phí điện tử   * Thông tin thồi gian thực * Hướng dẫn đỗ xe | Hệ thống định vị   * Hướng dẫn đỗ xe * Thông tin thồi gian thực   Thu phí điện tử   * Hệ thống bến đỗ * Hệ thống cảnh báo | * Thông tin thồi gian thực   Hệ thống tương tác  Vé điện tử  Camera giám sát |

Bảng 2.1: Dự đoán mật độ giao thông với hệ thống giao thông thông minh

Dấu ✓ thể hiện hệ thống được tích hợp với chức năng Dự đoán mật độ giao thông

### Hệ thống quản lý giao thông thông minh

Chính phủ và các cơ quan phụ trách giao thông chịu trách nhiệm về việc cung cấp các cơ sở hạ tầng và cơ sở hạ tầng hệ thống liên quan đến việc tăng cường an toàn giao thông. Hệ thống đường cao tốc và đường thứ cấp nói chung là khác nhau. Ví dụ về các cơ sở hạ tầng liên quan đến hệ thống ITS là:

* Giới hạn tốc độ: Các bảng hiệu thông báo chủ yếu được sử dụng để cảnh báo về giới hạn tốc độ được áp dụng trên một con đường. Mục đích là để giảm tốc độ trước khi tắc nghẽn xuất hiện, điều này sẽ dẫn đến một dòng chảy giao thông đồng nhất và hiệu quả hơn.
* Quản lý làn đường: làn dành riêng cho xe tải, xe buýt thường được sử dụng ở các nước phát triển để giảm lượng phương tiện lưu thông trên các hệ thống đường bộ vốn chỉ dành riêng cho xe máy và ô tô.
* Quản lý sự cố: Các sự cố xảy ra khi tham gia giao thông sẽ tác động tiêu cực đối với việc xử lý lưu lượng giao thông của một con đường. Thủ tục xử lý sự cố tốt hơn sẽ có thể hạn chế mật độ xảy ra sự cố trên các tuyến đường. Hơn nữa, dự báo nguy cơ xảy ra sự cố cũng sẽ giúp làm sáng tỏ vụ việc nhanh hơn.
* Hệ thống Cảnh báo: Các hệ thống này có thể cung cấp một số loại thông tin (như sương mù, ùn tắc, tai nạn, v.v.). Có một số hệ thống thông tin (dựa trên GPS) có thể được sử dụng cho sự phát triển của hệ thống cảnh báo.
* Camera giám sát: Closed Circuit Television (CCTV) máy quay ghi lại video hoặc ảnh chụp các tình huống được xác định trước đó. Mục tiêu chung là các luồng giao thông sẽ được sử dụng để phân tích các tài liệu video tự động và đưa ra các kết quả phân tích giao thông.
* Tự động nhận dạng xe (AVI): Xu hướng chung là nhận dạng xe tự động được thực hiện bằng thẻ để thực thi luật giao thông.
* Quản lý tốc độ thông minh (ISA): là một tên gọi chung cho các hệ thống, trong đó tốc độ của một chiếc xe được theo dõi thường xuyên trong một khu vực nhất định. Khi xe vượt quá giới hạn tốc độ, tốc độ được tự động điều chỉnh hoặc một cảnh báo về tốc độ sẽ được gửi đến cho người lái xe.
* Quản lý tải tọng (WIM): hỗ trợ công nghệ để kiểm tra xe tải về tải trọng của xe xem có vượt quá tải trọng cho phép không.
* Kiểm soát giao thông: Bộ điều khiển giao thông được sử dụng để điều tiết luồng giao thông tại nút giao thông.
* Thu phí tự động (ETC): Được thực hiện tại một hoặc nhiều điểm trên lộ trình giao thông, mục đích là để thu lệ phí một cách tự động. Làn đường và mức giá lý tưởng của làn đường được thực hiện bằng cách sử dụng một hệ thống tự động.
* Hệ thống thông tin thời gian thực: Hệ thống thông tin thời gian thực sử dụng dữ liệu thu thập bởi các trung tâm quản lý giao thông để thông báo cho người sử dụng đường của sự cố, sự chậm trễ v.v.
* Hướng dẫn đậu xe: hệ thống hướng dẫn đậu xe, dựa trên hệ thống định vị, cung cấp trình điều khiển với các thông tin liên về các chỗ trống trong bãi đậu xe.

### Hệ thống thông tin hành khách thông minh

Ngành công nghiệp xe hơi thế giới đã được bổ sung thêm một số hệ thống giao thông thông minh (ITS) trong các xe riêng. Các hệ thống tập trung vào an toàn, thực thi và kiểm soát, tính di động và hiệu quả, trước và trên chuyến đi. Các hệ thống này cũng hỗ trợ trong việc thực hiện các hệ thống bán vé và giá cả. Việc này không nhằm mục đích cố gắng để có một danh sách đầy đủ của hệ thống, mà nó được phát triển và đưa vào thị trường với các tính năng đầy đủ nêu trên để có thể giúp người dùng có thể tham gia giao thông thuận tiện hơn.

Hệ thống được tích hợp trong phương tiện cá nhân sẽ bao gồm các hệ thống sau:

* Hệ thống Định vị: hệ thống điện tử, cung cấp thông tin về các tuyến đường tới các tài xế.
* Hệ thống kiểm soát hành trình (CC): Hệ thống điều khiến đảm bảo tốc độ không vượt quá giới hạn.
* Hệ thống Cảnh báo: Hệ thống cảnh báo liên quan đến xe bao gồm hệ thống chống va chạm (sử dụng bộ cảm biến), các hệ thống thời tiết (thông qua đài phát thanh, hệ thống định vị v.v.), cảnh báo ùn tắc, tiết kiệm nhiên liệu (còn gọi là econometer), bẫy tốc độ cảnh báo (thông qua màn hình LCD hiển thị hoặc âm thanh), v.v.
* Hệ thống hộp đen: Những hệ thống này đã được sử dụng trong ngành hàng không trong nhiều năm, nhưng trong môi trường đường bộ là ý tưởng mới. Mục đích là để phân tích tình trạng xe và cảnh báo lái xe nếu vấn đề có thể xảy ra. Một ứng dụng có thể được phân tích tình trạng của chiếc xe tải để có thể cảnh báo lái xe nếu hành vi hệ thống phân tích chỉ ra rằng người lái xe đang ngủ, hoặc đang trong trạng thái thể chất không tốt để có thể lái xe.
* Tự động nhận dạng xe (AVI): Xu hướng chung là nhận dạng xe tự động được thực hiện bằng thẻ. Các tính năng nhân dạng của một thẻ có thể giúp việc thực thi luật giao thông.
* Hệ thống Bến đỗ: cảm biến được sử dụng để đo khoảng cách của một chiếc xe với các xe khác hoặc các đối tượng khác trong bãi đậu xe. Sử dụng hệ thống Bến đỗ, người dùng có thể đậu xe của mình chính xác hơn. Hơn nữa, có thể đậu xe một cách an toàn hơn trong những không gian nhỏ.
* Cảnh báo khoảng cách: Bảng cảm biến được sử dụng để cảnh báo trình điều khiển xe mà họ đang nhận được quá gần với các loại xe khác. Hệ thống sẽ cảnh báo khoảng cách và tốc độ xe cho tài xế.

Hệ thống thông tin hành khách thông minh là hệ thống hướng dẫn đậu xe, ISA, hệ thống thông tin thời gian thực, thu phí tự động ETC (bao gồm cả giá đường).

### Hệ thống giao thông công cộng thông minh

Một số hệ thống đã hoặc đang được phát triển để tăng cường giao thông công cộng. Những hệ thống này bao gồm:

* Quản lý theo nhóm: Các hệ thống dựa trên công nghệ định vị và liên kết phản hồi với những hoạt động đã được định nghĩa trước đó. Các nhà điều hành sẽ có thể theo dõi các phương tiện, phân tích hành vi của lái xe và thực hiện các bước nếu hành vi của lái xe là không đạt yêu cầu.
* Bán vé điện tử: Bán vé điện tử sẽ nâng cao hiệu quả của một hệ thống giao thông công cộng (thanh toán được thực hiện nhanh hơn) và sẽ cung cấp một môi trường an toàn hơn (ID người dùng có sẵn và nó sẽ khó khăn hơn cho bọn tội phạm để lại vô danh). Hơn nữa, từ một điểm kỹ thuật giao thông của xe, vé điện tử cung cấp cơ hội để cải thiện việc thu thập dữ liệu nhu cầu đi lại.
* Hệ thống vận chuyển cao tốc: Đây là hệ thống dẫn đường có khả năng duy trì tốc độ hoạt động vượt quá 200 km/h. Ví dụ: tàu cao tốc đầu tiên được sử dụng tại Nhật Bản khoảng 30 năm trước đây. Sau đó phương tiện cao tốc này đã được ứng dụng rộng rãi ở các nước Châu Âu nhờ lợi ích vượt trội của nó.
* Ưu tiên giao thông công cộng (PTP): Các hệ thống giảm thiểu các tác động tiêu cực của đèn giao thông cho giao thông công cộng. Nhiều hệ thống kiểm soát giao thông, như phân đường, theo dõi chu kỳ và hệ thống tối ưu hoá di chuyển có thể ưu tiên giao thông công cộng tại các nút giao. Các phương tiện giao thông công cộng cũng có thể được xác định (ví dụ qua thẻ).
* Thông tin thời gian thực: Liên quan đến giao thông công cộng, thông tin thời gian thực có thể được sử dụng để quản lý nhu cầu đi lại của người dân (Travel Demand Management - TDM). TDM tìm cách gây tác động đến hành vi của con người và khuyến khích một sự thay đổi từ tư nhân đến giao thông công cộng. Hệ thống ITS hỗ trợ để tạo ra sự thay đổi này.
* Hệ thống tích hợp: Để tăng cường sự hấp dẫn của giao thông công cộng so với các xe tư nhân (ví dụ, thời gian du lịch, thời gian chờ đợi tại các điểm dừng, v.v.) phải được cải thiện. Tài xế có được thông tin về giao thông công cộng thông qua việc đăng ký dịch vụ, sẽ giúp cải thiện thời gian đi lại. Tích hợp hệ thống sẽ làm giảm thời gian chờ đợi.

## Lợi ích của Hệ thống giao thông thông minh

Để thu thập được hiệu quả của các biện pháp ITS, dữ liệu từ các nghiên cứu khác nhau đã được thu thập và so sánh.

Các kết quả của những nghiên cứu này đã được thu thập được bằng cách sử dụng linh động các mô hình khác nhau. Hơn nữa, một số nghiên cứu chỉ ra nhu cầu của việc sử dụng Hệ thống giao thông thông minh sẽ ngày càng tăng lên trong tương lai.

Ngoài ra, thời gian nghiên cứu của các nghiên cứu khác nhau, nên đã có những ảnh hưởng nhất định đến kết quả. Mặc dù có những khác biệt, một kết quả chung về tác động của các biện pháp ITS đã được tìm ra.

Trong một nghiên cứu được thực hiện bởi Ludmann và cộng sự vào năm 1997 [8], 100% số xe đã được trang bị hệ thống hành trình. Ước tính, sử dụng mô hình Pelops, đã được thực hiện cho đường cao tốc cũng như giao thông đô thị. Hai hệ thống khác nhau đã được thử nghiệm (lần thử nghiệm thứ hai đã trôi chảy). Những thay đổi về tốc độ và lưu lượng giao thông được tính toán. Tốc độ trung bình trong kịch bản đầu tiên giảm 13%, trong khi tốc độ tăng 6% trong lần thử nghiệm thứ hai. Cả hai kịch bản cho thấy sự gia tăng trong kết quả đầu ra (12% -14%).

Hồ sơ động được áp dụng cho đường cao tốc. Tampere và các cộng sự đã tìm thấy một sự gia tăng đáng kể trong sức chứa của các con đường (chỉ ra bởi các kết quả thu được). Trong nghiên cứu này, ba làn đường cao tốc truyền thống đã được thay thế bởi bốn làn xe nhỏ hơn với tốc độ tối đa, đã khiến cho sự tăc nghẽn xảy ra thấp hơn trong giờ cao điểm. Sự gia tăng công suất dự kiến 20% là rất hứa hẹn. Các nghiên cứu được thực hiện bởi Stemerding và các cộng sự vào năm 1990 [9, chương 1] đã chỉ ra rằng sự gia tăng lưu lượng giao thông tổng thể nằm trong khoảng 5%. Trong nghiên cứu này, tốc độ tối đa giảm từ 100 km/h đến 70 km/h. Sự giảm (4%) trong số các điểm dừng là một dấu hiệu cho thấy tình hình an toàn giao thông đã được cải thiện. Nói chung, thời gian đi lại đã được giảm (lên đến 41%). Trong một trường hợp khác, thời gian đi lại tăng 16%.

Điều khiển động thông tin tuyến đường là Hệ thống thông báo hiệu (Variable Message Signs - VMS), thông báo cho người lái xe về tắc nghẽn trước, chủ yếu là trên đường cao tốc và thời gian đi dự kiến. Các nghiên cứu của Van Straaten vào năm 2001 cho thấy rằng các loại Hệ thống báo hiểu đã giúp giảm mức độ nghiêm trọng của tình trạng tắc nghẽn. Thời gian di chuyển trung bình giảm lên đến 42%. Kết quả này là đáng kể. Mặc dù không rõ là bao nhiêu người tham gia giao thông làm theo các gợi ý, nhưng đã ít tuyến đường bị tắc nghẽn, nghiên cứu này chỉ ra rằng tỷ lệ phần trăm là đủ cao để tạo sự khác biệt.

Những tác động của một hệ thống quản lý đường cao tốc là rất hứa hẹn, việc giảm lượng phương tiện tham gia giao thông trong khoảng thời gian dự kiến lên đến 48% là rất đáng kể. Theo thống kê của Thomas vào năm 2001 [10] đã chỉ ra Sự gia tăng ước tính công suất của đường cao tốc (thông lượng) lên đến 25%, đây là một chỉ số đáng chú ý. Kết quả xuất hiện rất tích cực vì một hệ thống quản lý đường cao tốc là một sự kết hợp các biện pháp:

* Thông báo tín hiệu.
* Nâng cao hệ thống thông tin di động, chẳng hạn như trong xe, giám sát.
* Thu phí tự động hoặc thanh toán tiền vé điện tử.
* Camera giám sát (quản lý sự cố) và nhận dạng xe.
* Báo cáo Radio, tuần tra trên không và như vậy.

Câu hỏi đặt ra là làm thế nào để cân đối nỗ lực và chi phí để cung cấp các gói này một cách hợp lý.

Làn xe chở khách (High Occupancy Vehicle - HOV) được tạo ra trên đường cao tốc, tập trung vào một sự thay đổi trong thời gian đi lại. Theo nghiên cứu của Dahlgen vào năm 1998 và Johnston vào năm 1996 đã cùng chỉ ra thời gian đi lại ước tính giảm lên đến 8% khi có một làn xe chở khách riêng trên đường cao tốc. Tuy nhiên, trong nhiều trường hợp, thời gian đi lại, đặc biệt là cho người không đi làn HOV, tăng lên. Trong nghiên cứu được tiến hành bởi Johnston vào năm 1996 sự gia tăng này là lên đến 200%.

Đồng nhất giới hạn tốc độ luôn luôn dẫn đến giao thông đồng nhất hơn trên một hệ thống đường cao tốc.

Mục đích chính của hệ thống Thích ứng tốc độ thông minh (Intelligent Speed Adaptation - ISA) là cải thiện an toàn đường bộ. Vì nó giúp lái xe không vượt quá tốc độ giới hạn trong hệ thống bắt buộc theo mô hình, tốc độ trung bình của xe giảm (34% của xe vượt quá giới hạn tốc độ trước khi sự ra đời của ISA). Các mô hình đã được thực hiện đối với một mạng lưới các đường cao tốc và một số giới hạn của đường phụ.

Làn trả phí là làn xe chuyên dụng trên đường cao tốc, nơi tập hợp các điểm thu phí. Quá trình thu phí chủ yếu sử dụng các thiết bị điện tử. Các tác dụng ước tính của làn nói chung là tích cực. Nghiên cứu của Schoemakers và các cộng sự vào năm 2000 chỉ ra thời gian đi lại nhìn chung đã giảm 33%. Các trình thiết bị điều khiển không tăng thời gian đi lại của phương tiện.

Làn xe cao tốc là làn đường trên của một đường cao tốc mà chỉ mở cửa cho giao thông trong giờ cao điểm. Chúng thường được sử dụng trong các hướng khác nhau trong suốt buổi sáng và buổi tối cao điểm. Làn cao tốc giảm nguy cơ tắc nghẽn và giữ cho dòng di chuyển đồng nhất hơn. Các nghiên cứu của Stemerding đã chỉ ra Các thông số tăng khoảng 5% trong giờ cao điểm.Tình hình an toàn giao thông được cải thiện đôi chút. WESTRA và Bosch cũng điều tra những ảnh hưởng của làn xe cao điểm. Họ nhận thấy tổng thời gian đi lại trên toàn bộ mạng lưới giao thông đã giảm khoảng 21%. Tuy nhiên, họ cũng chỉ ra rằng tổng thời gian đi lại trên một số các bộ phận của mạng lưới giao thông tăng 40% trong thời gian du lịch. Nghiên cứu của Bosch vào năm 2003 tìm thấy một thông mà thay đổi từ -5% đến + 6%. Nghiên cứu này xác định nguy cơ cao tăng thời gian di chuyển (giữa -9% đến + 50%).

Nói chung có thể kết luận rằng biện pháp ITS cải thiện rõ việc quản lý luồng giao thông trên đường cao tốc và thường tạo ra môi trường an toàn hơn.

## Vai trò của Dự đoán mật độ giao thông trong Hệ thống giao thông thông minh

Trong Hệ thống giao thông thông minh (ITS), việc dự đoán mật độ giao thông đóng một vai trò rất quan trọng. Các kết quả dự đoán mật độ giao thông được ITS sử dụng cho các chức năng:

* Giới hạn tốc độ: hệ thống sẽ dựa vào mật độ giao thông được dự đoán trên một khung đường và một khung giờ để định ra tốc độ tối đa lưu thông trên khung đường và khung giờ đó là an toàn nhất.
* Quản lý làn đường: đối với các hệ thống giao thông tiên tiến, làn đường thường đường chia làm nhiều làn nhỏ và việc phân bố làn đường sẽ được thay đổi thích hợp.
* Hệ thống cảnh báo: khi có dự đoán một khung đường sẽ xảy ra ùn tắc, hệ thống sẽ thực hiện gửi tin nhắn cánh bảo cho người tham gia giao thông để họ có lựa chọn đường đi phù hợp.
* Kiểm soát hành trình: việc dự đoán được mật độ giao thông sẽ giúp hệ thống xây dựng được một hành trình phù hợp cho người tham gia giao thông, giúp luồng giao thông được trôi chảy hơn.

Dự đoán mật độ giao thông là tiền đề để xây dựng nên các chức năng khác. Một trong những ưu điểm của việc dự đoán là tính chủ động. Hệ thống và người tham gia giao thông có thể chủ động đưa ra các định hướng di chuyển nhằm mục đích chung là việc giao thông được thuận lợi.

# Chương 3: Xây dựng mô hình dự đoán mật độ giao thông

Phương pháp dự báo tình trạng giao thông ngắn hạn có thể được giải quyết bằng một số giải pháp như sau:

+ Phương pháp thống kê: dựa vào thông tin về trạng thái giao thông của ngày hôm trước để dự báo cho ngày hôm sau. Phương pháp này có ưu điểm là đơn giản, dựa trên thực tế là trong các giờ làm việc là cố định không thay đổi nên thông thường tình trạng giao thông là ổn định theo các khung giờ, ngoại trừ một số trường hợp bất thường xảy ra (như có sự kiện được tổ chức có thể dẫn đến cấm một số tuyến phố, dẫn đến một số tuyến khác phải gánh thêm lưu lượng của người tham gia giao thông trên các tuyến bị cấm). Nhưng phương pháp này có một số nhược điểm là không quản lý được một số tham số động như ngày cuối tuần (một số lượng lớn người sẽ được nghỉ làm và sẽ thay đổi lưu lượng giao thông), ngày nghỉ (một số lượng lớn người có thể về quê nên sẽ có một số tuyến sẽ tăng lưu lượng, sau đó lưu lượng giao thông nội thành sẽ giảm trong những ngày nghỉ, và đến hết giai đoạn nghỉ thì sẽ có một số tuyến tăng lưu lượng vì người đi làm quay lại thủ đô làm việc).

+ Phương pháp dùng luật: xây dựng một số luật kết hợp với thông tin thống kê để ước lượng lưu lượng. Ví dụ, dùng thông tin thống kê của ngày làm việc làm tham số ước lượng cho ngày làm việc, ngày nghỉ ước lượng cho ngày nghỉ, … Ưu điểm của phương pháp này cũng là đơn giản, tuy nhiên nó vẫn có nhược điểm là phương pháp tĩnh không quản lý được một số tham số động như thời tiết, hay sự ùn tắc cục bộ của một số tuyến đường liên quan.

+ Phương pháp dùng học máy (Machine learning): có rất nhiều phương pháp học máy, tuy nhiên có một lớp giải thuật có thể ứng dụng cho bài toán dự đoán lưu lượng giao thông là các giải thuật phân lớp. Khi ta xác định một số mức của lưu lượng như: cấp 1 (tắc nghẽn), mức 2 (rất đông, vận tốc di chuyển chậm < 12km/h), mức 3 (đường đông, vận tốc di chuyển vừa phải từ 12km/h – 25km/h), mức 4 (đường thoáng, tốc độ di chuyển từ 25km/h đến 35km/h), mức 5 (đường rất thoáng tốc độ di chuyển > 35km/h), thì ta có thể xây dựng một bộ phân lớp để xác định trạng thái giao thông của một tuyến đường thuộc vào mức nào nếu ta coi mỗi mức tương ứng với một lớp. Ưu điểm của phương pháp này là động, nó thay đổi theo tình trạng thực tế của các tham số đầu vào, do vậy trong luận văn này, tác giả lựa chọn đi theo hướng này.

Việc dự đoán mật độ giao thông cần phải dựa vào nhiều nguồn thông tin khác nhau. Mỗi nguồn thông tin là một điều kiện cũng như một căn cứ để từ đó xác định ra mật độ giao thông tại một thời điểm. Để có thể có được dự đoán chính xác nhất, chúng ta cần phải căn cứ các thông tin đã có (ngày, giờ, cung đường gì ,…) để từ đó đưa ra được dự đoán chính xác nhất. Khi có được một tập hợp các thông tin là chúng ta đã có được một tập dữ liệu, để có thể từ tập dữ liệu này phân tích và đưa ra các kết quả dự đoán ta có thể sử dụng mô hình phân lớp dữ liệu. Quá trình phân lớp dữ liệu là quá trình mà từ một tập dữ liệu mẫu có sẵn, hệ thống sẽ phân tích, tìm ra các thuộc tính tương đồng để xây dựng mô hình và đưa ra các kết quả phân lớp chính xác. Kết quả phân lớp ở đây sẽ chính là mật độ giao thông mà chúng ta đang hướng đến.

## Bài toán phân lớp dữ liệu

Là quá trình phân lớp một đối tượng dữ liệu vào một hay nhiều lớp cho trước nhờ một mô hình phân lớp mà mô hình này được xây dựng dựa trên một tập hợp các đối tượng dữ liệu đã được gán nhãn từ trước gọi là tập dữ liệu học (tập huấn luyện).

Quá trình phân lớp còn được gọi là quá trình gán nhãn cho các đối tượng dữ liệu.Như vậy, phân lớp cũng là tiên đoán lại lớp của nhãn.

Có nhiều bài toán phân lớp dữ liệu, như phân lớp nhị phân, phân lớp đa lớp, phân lớp đa trị,…. Phân lớp nhị phân là quá trình tiến hành việc phân lớp dữ liệu vào một trong hai lớp khác nhau dựa vào việc dữ liệu đó có hay không một số đặc tính theo quy định của bộ phân lớp. Phân lớp đa lớp là quá trình phân lớp với số lượng lớp lớn hơn hai.

Trong phân lớp đa trị, mỗi đối tượng dữ liệu trong tập huấn luyện cũng như các đối tượng mới sau khi được phân lớp có thể thuộc vào từ hai lớp trở lên.

Với ví dụ là bài toán Dự đoán mật độ Giao thông. Mỗi một đối tượng dữ liệu trong tập huấn luyện là một trường hợp giao thông với một số điều kiện nhất định. Các dữ liệu huấn luyện sẽ không đơn giản chỉ là thông tin tại một thời điểm mà một bộ huấn luyện sẽ ra rất nhiều thời điểm khác nhau. Một lớp chính là một tập hợp các dữ liệu được đánh giá theo giá trị bao gồm từ 1 đến 5 tương ứng với mật độ giao thông từ thưa thớt cho tới đông đúc. Mỗi một dữ liệu phân lớp sẽ có các giá trị khác nhau, dựa vào các giá trị này , áp dụng các thuật toán phân lớp sẽ phân tích dữ liệu đầu vào và phân các giá trị đó vào các lớp tương ứng.

Việc dự đoán mật độ giao thông có 3 hướng tiếp cận chính là thông qua quy luật, mô hình hóa và học máy. Trong phương pháp tiếp cận thì học máy có nhiều ưu điểm như không mất thời gian đưa ra các luật, học từ dữ liệu huấn luyện, dễ dàng mở rộng và tái cấu trúc. Các bộ phân loại thường sử dụng trong phương pháp học máy là Support Vector Machine(SVM), Naive Bayes, J48, Neural Network,Maximum Entropy, Decision Tree, Nearest-Neighbors, Sparse Network of Winnows(SNoW).

Trong luận văn nghiên cứu này, tôi sẽ sử dụng mô hình chính là Decision Tree để học dữ liệu mẫu và đưa ra các dự đoán về mật độ giao thông từ các dữ liệu mẫu.

## Mô hình dự đoán mật độ giao thông

Quá trình phân lớp dữ liệu thường gồm hai bước: Xây dựng mô hình và Sử dụng mô hình

### Mô hình đề xuất

**Mục tiêu**: xây dựng một mô hình mô tả một tập các lớp dữ liệu hay các khái niệm định trước.

Một mô hình sẽ được xây dựng dựa trên việc phân tích các đối tượng dữ liệu đã được gán nhãn từ trước. Tập các mẫu dữ liệu này còn được gọi là tập dữ liệu huấn luyện (training data set).

Ta có mô hình áp dụng với bài toán mật độ Giao thông



Hình 1:Mô hình dự đoán mật độ giao thông

### Sử dụng mô hình

Sử mô hình đã xây dựng ở bước trước để phân lớp dữ liệu mới.

Trong mô hình phân lớp, thuật toán phân lớp giữ vai trò trung tâm, quyết định tới sự thành công của mô hình phân lớp. Do vậy chìa khóa của vấn đề phân lớp dữ liệu là tìm ra được một thuật toán phân lớp nhanh, hiệu quả, có độ chính xác cao và có khả năng mở rộng được. Trong đó khả năng mở rộng được của thuật toán được đặc biệt trú trọng và phát triển.

Một số kỹ thuật phân lớp được sử dụng và phát triển với đề tại Dự đoán mật độ giao thông:

* Kỹ thuật phân lớp Super Vector Machine(SVM)
* Kỹ thuật phân lớp Naïve Bayes
* Kỹ thuật phân lớp J48 (Decision Tree)
* Kỹ thuật phân lớp Neural Network

Với kỹ thuật phân lớp SVM kernel sẽ được sử dụng trong mô hình là Polynomial Kernel. Trong học máy, Polynomial Kernel là một hàm kernel được SVM sử dụng, nó thể hiện sự tương đồng của các vector (mẫu huấn luyện) trong không gian đặc trưng trên đa thức của các biến và cho phép học theo mô hình phi tuyết tính.

Mỗi kỹ thuật phân lớp có ưu điểm và đặc điểm riêng. Độ chính xác của các kỹ thuật cũng khác nhau.

## Các bước xây dựng mô hình Dự đoán mật độ giao thông

### Bước 1: Tạo dữ liệu thực nghiệm

#### Lựa chọn đặc tính để sinh dữ liệu

Do số liệu cụ thể liên quan đến các đoạn đường, thời điểm tắc đường thường không được công bố nên chúng ta không có số liệu thật để kiểm thử chương trình.

Tuy nhiên, như đã trình bày ở các chương trước đó, dữ liệu liên quan đến địa điểm tắc đường thường có tính quy luật, nên dựa vào các quy luật này ta có thể sinh ra bộ dữ liệu sát với số liệu thật.

Cụ thể các thông tin được dựa vào để sinh số liệu bao gồm:

* Các trường hợp tắc đường thường xuất hiện ở trên những cung đường nhất định, vậy nên **Tên đường** chính là 1 thông tin để xác định tình hình mật độ giao thông.
* Các trường hợp tắc đường thường xuất hiện trong những ngày nhất định trong nằm (ngày lễ, ngày nghỉ, …) vậy nên thời gian **Ngày/Tháng** là 1 thông tin để xác định mật độ giao thông.
* Trong một ngày, mật độ giao thông ở các khoảng thời gian là khác nhau, nhưng tình trạng tắc đường thường chỉ xuất hiện ở những khung giờ nhất định, vậy nên **Giờ** là 1 thông tin để xác định mật độ giao thông.
* Trong một tuần, mật độ giao thông ở các ngày trong tuần là khác nhau, thứ 2 đầu tuần sẽ đông hơn và các ngày trong tuần sẽ có mật độ giao thông cao hơn các ngày thứ 7 và chủ nhật. Vậy nên **Ngày trong tuần** là 1 thông tin để xác định mật độ giao thông
* Mật độ giao thông trên 1 cung đường thường giống nhau ở cùng khoảng thời gian, ví dụ khoảng 5h sáng ở cung đường A mật độ giao thông thường ở mức thấp. Tuy nhiên trong một số trường hợp đột biến phát sinh (do thời tiết, do sửa chữa, do tai nạn giao thông,…) mà mật độ thỉnh thoảng sẽ có sự đột biến khác với những ngày còn lại. Tỉ lệ phát sinh đột biến được lựa chọn là 20%. Như vậy việc sinh mật độ giao thông tại 1 thời điểm trên 1 cung đường sẽ có tỉ lệ 20% đột biến so với các ngày còn lại.

Như vậy, các Đặc tính (Attribute) được lựa chọn để xây dựng bộ kiểm thử bao gồm:

* Tên cung đường (RoadId)
* Ngày (Date)
* Tháng (Month)
* Giờ (Hour)
* Ngày trong tuần (Day)

Trong đó Trạng thái giao thông (state) chính là mục tiêu phân lớp mà bài toán Dự đoán mật độ giao thông hướng đến.

#### Thực hiện sinh dữ liệu

Phần mềm Excel được sử dụng để xây dựng bộ dữ liệu và số class (mật độ tắc đường) được lựa chọn ở đây là 6.

Dựa vào các đặc tính được lựa chọn để sinh dữ liệu, các bước để sinh dữ liệu bao gồm:

* Sinh thông tin giờ: một ngày có 24 giờ (0h -> 23h), thông tin dữ liệu của 1 ngày bất kỳ sẽ phải có thông tin của đẩy đủ 24 giờ. Tức là mỗi một ngày sẽ phải có 24 bản ghi dữ liệu tương ứng với trạng thái của 24h.
* Sinh thông tin ngày: một tháng có từ 30->31 ngày, thông tin của 1 tháng bất kỳ sẽ phải có đầy đủ thông tin của các ngày trong tháng đó. Do đó sẽ cần phải sinh thông tin cho khoảng 31 x 24 = 744 bản ghi dữ liệu. Như vậy mỗi tháng sẽ phải có khoảng 744 bản ghi đại diện cho mật độ của đường tại 744 thời điểm.
* Sinh thông tháng: 1 năm có 12 tháng, như vậy, thông tin của 1 năm sẽ có khoảng 12 x 744 = 8928 bản ghi đại diện cho mật độ của 8928 thời điểm trong năm.
* Sinh thông tin đường: do ta sẽ sinh thông tin đại diện của mật độ tắc đường trong 1 năm nên mỗi một đường sẽ cần sinh khoảng 8928 bản ghi đại diện cho các thời điểm tắc đường trong năm. Giả sử ta có N cung đường, thì bộ dữ liệu sinh sẽ cần có khoảng N x 8928 bản ghi đại diện cho các thời điểm của N cung đường trong năm
* Sinh thông tin mật độ giao thông: mật độ giao thông dựa vào khoảng thời gian trong ngày để đánh giá, do đó một ngày sẽ được chia làm các khung giờ khác nhau để đánh mật độ giao thông. Cụ thể:
  + Từ 0h -> 5h độ tắc đường sẽ là 1 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 2 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=0,D2<=5),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(1,2),1) |

* + Từ 6h -> 8h độ tắc đường sẽ là 5 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 6 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=6,D2<=8),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(5,6),5) |

* + Từ 9h -> 10h độ tắc đường sẽ là 3 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 4 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=9,D2<=10),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(3,4),3) |

* + Từ 11h -> 12h độ tắc đường sẽ là 4 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 5 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=11,D2<=12),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(4,5),4) |

* + Từ 13h -> 16h độ tắc đường sẽ là 1 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 2 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=13,D2<=16),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(2,3),2) |

* + Từ 17h -> 19h độ tắc đường sẽ là 4 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 5 và 6 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=17,D2<=19),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(4,6),4) |

* + Từ 20h -> 22h độ tắc đường sẽ là 2 tuy nhiên khoảng 20% tỉ lệ phát sinh độ tắc đường là 3 do có sự cố bất thường. Công thức Excel sinh độ tắc đường khung giờ này như sau:

|  |
| --- |
| IF(AND(D2>=20,D2<=22),IF(RANDBETWEEN(1,5)=1,RANDBETWEEN(2,3),2 |

* + Các khoảng thời gian không nằm trong tính toán sẽ đặt là 1 và sẽ random khoảng 20% độ tắc là 2. Ta có công thức sinh mật độ tắc đường toàn bộ như sau:

|  |
| --- |
| ="S"& IF((IF(AND(E7>=0,E7<=5),IF(RANDBETWEEN(1,5)=1,2,1),  IF(AND(E7>=6,E7<=8),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,3),2),  IF(AND(E7>=9,E7<=10),IF(RANDBETWEEN(1,5)=1,4,5),  IF(AND(E7>=11,E7<=12),IF(RANDBETWEEN(1,5)=1,5,6),  IF(AND(E7>=13,E7<=16),IF(RANDBETWEEN(1,5)=1,5,4),  IF(AND(E7>=17,E7<=19),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,5),6),  IF(AND(E7>=20,E7<=22),IF(RANDBETWEEN(1,5)=1,3,2),IF(RANDBETWEEN(1,5)=1,2,1)))))))) + IF(ISNA(VLOOKUP(DATEVALUE(MID(B7,2,2)&"/"&C7),ngaynghi!$A$1:$B$9,2,FALSE)),0,VLOOKUP(DATEVALUE(MID(B7,2,2)&"/"&C7),ngaynghi!$A$1:$B$9,2,FALSE))) = 0, 1, IF((IF(AND(E7>=0,E7<=5),IF(RANDBETWEEN(1,5)=1,2,1),  IF(AND(E7>=6,E7<=8),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,3),2),  IF(AND(E7>=9,E7<=10),IF(RANDBETWEEN(1,5)=1,4,5),  IF(AND(E7>=11,E7<=12),IF(RANDBETWEEN(1,5)=1,5,6),  IF(AND(E7>=13,E7<=16),IF(RANDBETWEEN(1,5)=1,5,4),  IF(AND(E7>=17,E7<=19),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,5),6),  IF(AND(E7>=20,E7<=22),IF(RANDBETWEEN(1,5)=1,3,2),IF(RANDBETWEEN(1,5)=1,2,1)))))))) + IF(ISNA(VLOOKUP(DATEVALUE(MID(B7,2,2)&"/"&C7),ngaynghi!$A$1:$B$9,2,FALSE)),0,VLOOKUP(DATEVALUE(MID(B7,2,2)&"/"&C7),ngaynghi!$A$1:$B$9,2,FALSE))) = 7,6,IF(AND(E7>=0,E7<=5),IF(RANDBETWEEN(1,5)=1,2,1),  IF(AND(E7>=6,E7<=8),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,3),2),  IF(AND(E7>=9,E7<=10),IF(RANDBETWEEN(1,5)=1,4,5),  IF(AND(E7>=11,E7<=12),IF(RANDBETWEEN(1,5)=1,5,6),  IF(AND(E7>=13,E7<=16),IF(RANDBETWEEN(1,5)=1,5,4),  IF(AND(E7>=17,E7<=19),IF(RANDBETWEEN(1,5)=1,IF(D7="SU",2,5),6),  IF(AND(E7>=20,E7<=22),IF(RANDBETWEEN(1,5)=1,3,2),IF(RANDBETWEEN(1,5)=1,2,1)))))))))) |

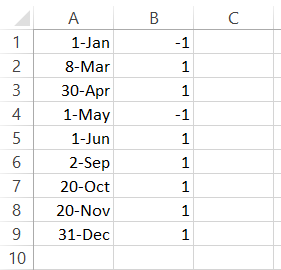
***Notes:*** Trong đó D2 là cell chứa thông tin giờ.

#### Thực hiện sinh dữ liệu với thông tin ngày nghỉ lễ

Ngoài các yếu tố như giờ giấc, ngày trong tuần, mật độ giao thông còn bị ảnh hưởng bởi một yếu tố nữa là ngày nghỉ. Tùy từng đặc trưng của riêng các ngày nghỉ lễ (lễ tết, noel, ngày nhà giáo Việt Nam, …) mà mật độ giao thông cũng sẽ có các sự biến đổi khác nhau. Để bám sát với thực tế hơn, dữ liệu thực nghiệm sẽ được bổ sung thêm thông tin ngày nghỉ. Trong luận văn này, ngoài dữ liệu thực nghiệm thông thường, tác giả sẽ thực nghiệm với dữ liệu có bổ sung thêm thông tin xét đến các ngày nghỉ lễ để có thể đưa ra được các đánh giá và so sánh chi tiết hơn khi áp dụng mô hình phân lớp vào dự đoán mật độ giao thông.

Trong một năm sẽ có rất nhiều ngày nghỉ lễ và các ngày nghỉ lễ này mật độ giao thông thường có tính quy luật. Ví dụ, ngày 1/1 hằng năm là ngày tết theo Dương lịch. Với ngày nghỉ này mọi người thường nghỉ ngơi đi du lịch hoặc về quê từ hôm trước, nên trên một số tuyến đường vào ngày này mật độ giao thông sẽ giảm mạnh. Ngày 8/3 là ngày quốc tế phụ nữ, đây là ngày lễ nhưng không phải ngày nghỉ, ngày này mọi người thường ra đường rất nhiều để đi ăn, liên hoan, tụ tập bạn bè nên mật độ giao thông trong ngày này trên một số tuyến đường sẽ tăng mạnh. Ngày 30/4 là ngày nghỉ đầu tiên trong đợt nghỉ lễ thống nhất đất nước, có nhiều thời gian nên 30/4 mọi người sẽ về quê hoặc bắt đầu đi du lich vào ngày này nên mật độ sẽ tăng. Như vậy, với quy luật có thể nhận thấy được, ta có thể tập hợp các thông tin lại và đưa ra một danh sách các ngày nghỉ lễ và biến động quy luật của nó. Dữ liệu ngày nghỉ này sẽ được đưa thêm vào điều kiện sinh dữ liệu thực nghiệm để tăng các tình huống xảy ra.

Để sinh dữ liệu với thông tin ngày nghỉ lễ, ta tạo một sheet trong file excel sinh dữ liệu thực nghiệm với thông tin là các ngày nghỉ lễ trong năm. Cụ thể, ta tạo thêm sheet [ngaynghi] với nội dung như sau:



Trong đó:

* Cột A: thông tin ngày nghỉ lễ
* Cột B: biến động mật độ giao thông trong ngày

Dựa vào bảng này, ta có thông tin quy luật biến động giao thông và các ngày nghỉ và ngày lẽ như sau:

* Ngày 1/1 mật độ giao thông sẽ giảm 1 so với bình thường
* Ngày 8/3 mật độ giao thông sẽ tăng 1 so với bình thường
* Ngày 30/4 mật độ giao thông sẽ tăng 1 so với bình thường

Như vậy khi tạo dữ liệu thực nghiệm trên excel, ta sẽ sử dụng công thức sau để lấy biến động mật độ giao thông:

|  |
| --- |
| IF(ISNA(VLOOKUP(DATEVALUE(MID(B5,2,2)&"/"&C5),ngaynghi!$A$1:$B$9,2,FALSE)),0,VLOOKUP(DATEVALUE(MID(B5,2,2)&"/"&C5),ngaynghi!$A$1:$B$9,2,FALSE)) |

Công thức tính trên thể hiện các tính mật độ thực nghiệm theo ngày. Trường hợp ngày thực nghiệm nằm trong danh sách nghỉ lễ trong sheet [ngaynghi] như đã xây dựng, kết quả trả về sẽ là 1 hoặc -1 tùy thuộc vào định nghĩa biến động mật độ giao thông trong sheet [ngaynghi]. Trường hợp ngày thực nghiệm không nằm trong danh sách ngày nghỉ lễ, biến động mật độ giao thông sẽ trả về 0. Kết hợp với công thức tính mật độ như ở mục b cùng chương, ta sẽ xây dựng được bộ dữ liệu có thêm tính chất ngày nghỉ lễ.

### Bước 2: Chia dữ liệu training và test

Phương tức test 70-30 là phương thức test mà Bộ dữ liệu huấn luyện được chia làm 2 phần:

* 70% dữ liệu được sử dụng để tranning
* 20% dữ liệu được sử dụng để test

Ta có các tập dữ liệu tranning và test như sau:

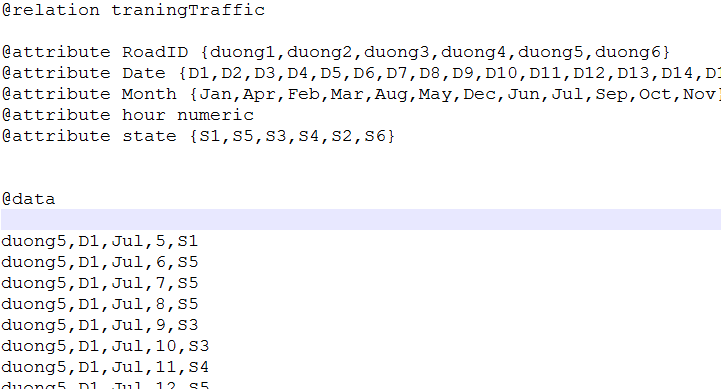
Tool sử dụng: Weka (Waikato Environment for Knowledge Analysis)

* Dữ liệu tranning:
  + Số lượng data: 53568
  + Số Attribute: 4
  + Số class: 6



Hình 2: Test 70-30 – Dữ liệu training

* Dữ liệu test:
  + Số lượng data test: 15000
  + Số Attribute: 4
  + Số class: 5

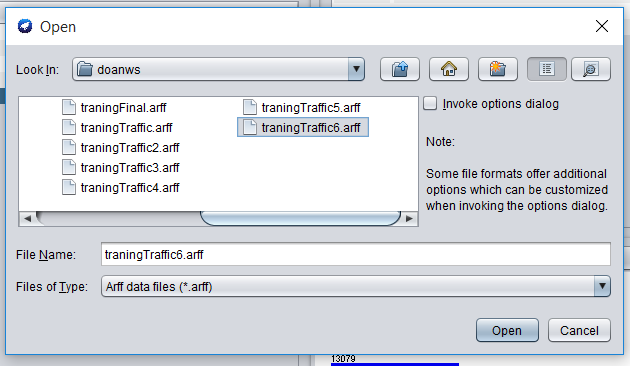


Hình 3: Test 70-30 – Dữ liệu Test

### Bước 3: Huấn luyện

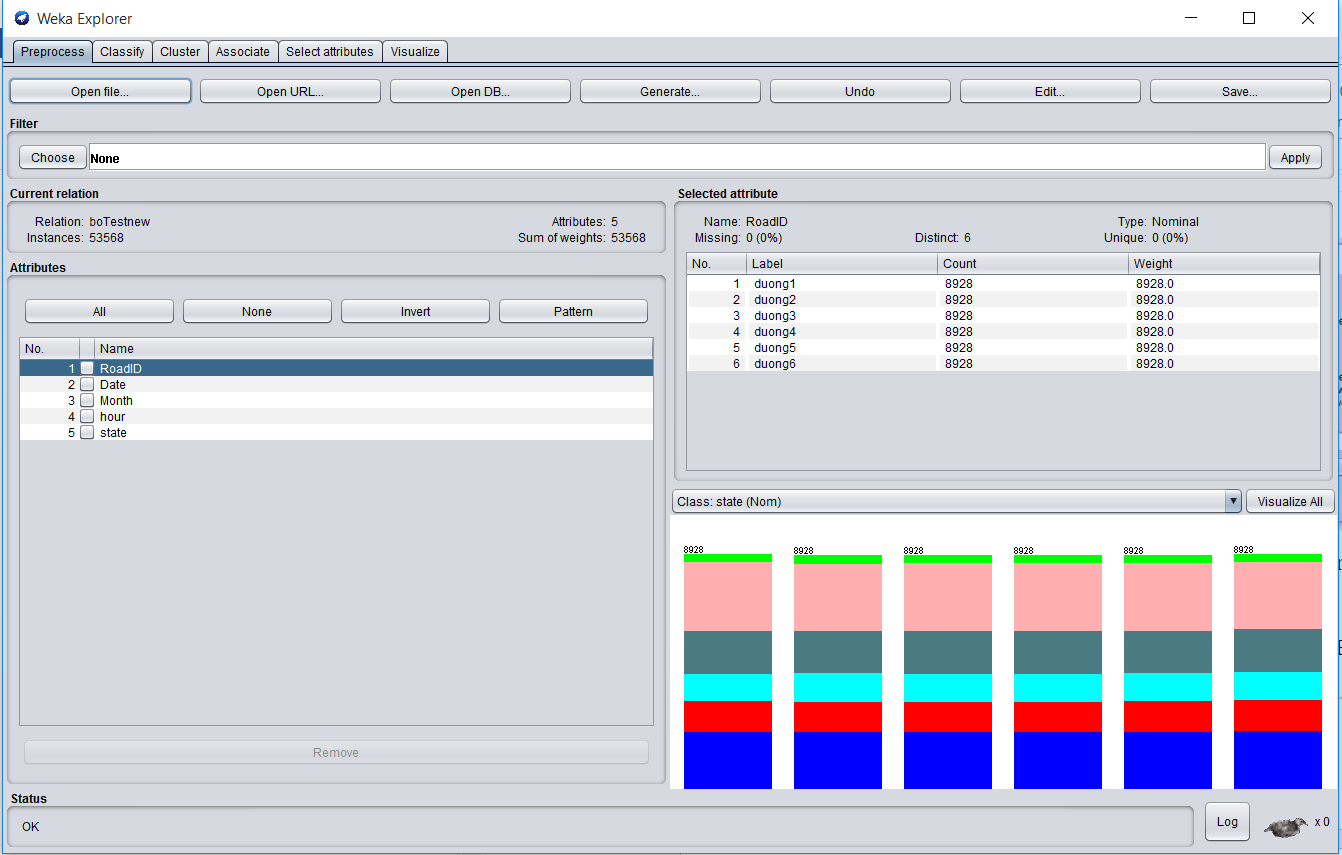
Sau khi có được dữ liệu thực nghiêm ta sẽ tiến hành huấn luyện dữ liệu trên công cụ Weka.

* Từ màn hình giao diện Weka Explorer ta chọn openfile và chọn file dữ liệu có định dạng arff được sử dụng để làm dữ liệu huấn luyện:



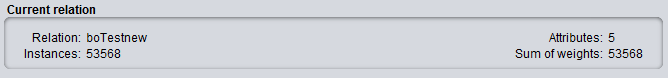
Hình 4: Lựa chọn bộ dữ liệu huấn luyện

* Sau khi chọn xong bộ dữ liệu huấn luyện, màn hình Process sẽ hiển thị đầy đủ thông tin của tập dữ liệu:



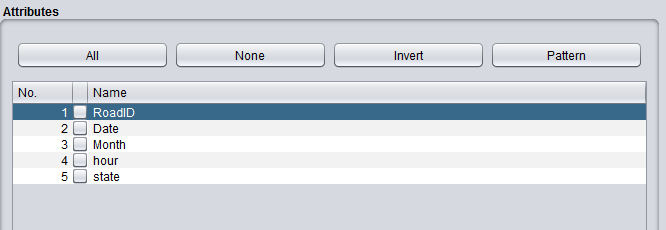
Hình 5: Thông tin dữ liệu huấn luyện

* Chi tiết các thông tin huấn luyện như sau:



Hình 6: Thông tin tập Quan hệ hiện tại

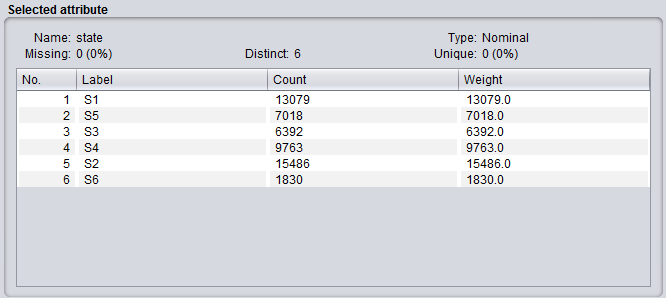
* + Tên tập quan hệ: boTestNew
  + Số thể hiện: 53568
  + Số Attributes: 5
* Thông tin chung các Attribute như sau:



Hình 7: Thông tin các Attributes

Có 5 Attribute: RoadId, Date, Month, Hour, State

* Thông tin chi tiết của các Attribute như sau:



Hình 8: Thông tin chi tiết của các Attributes

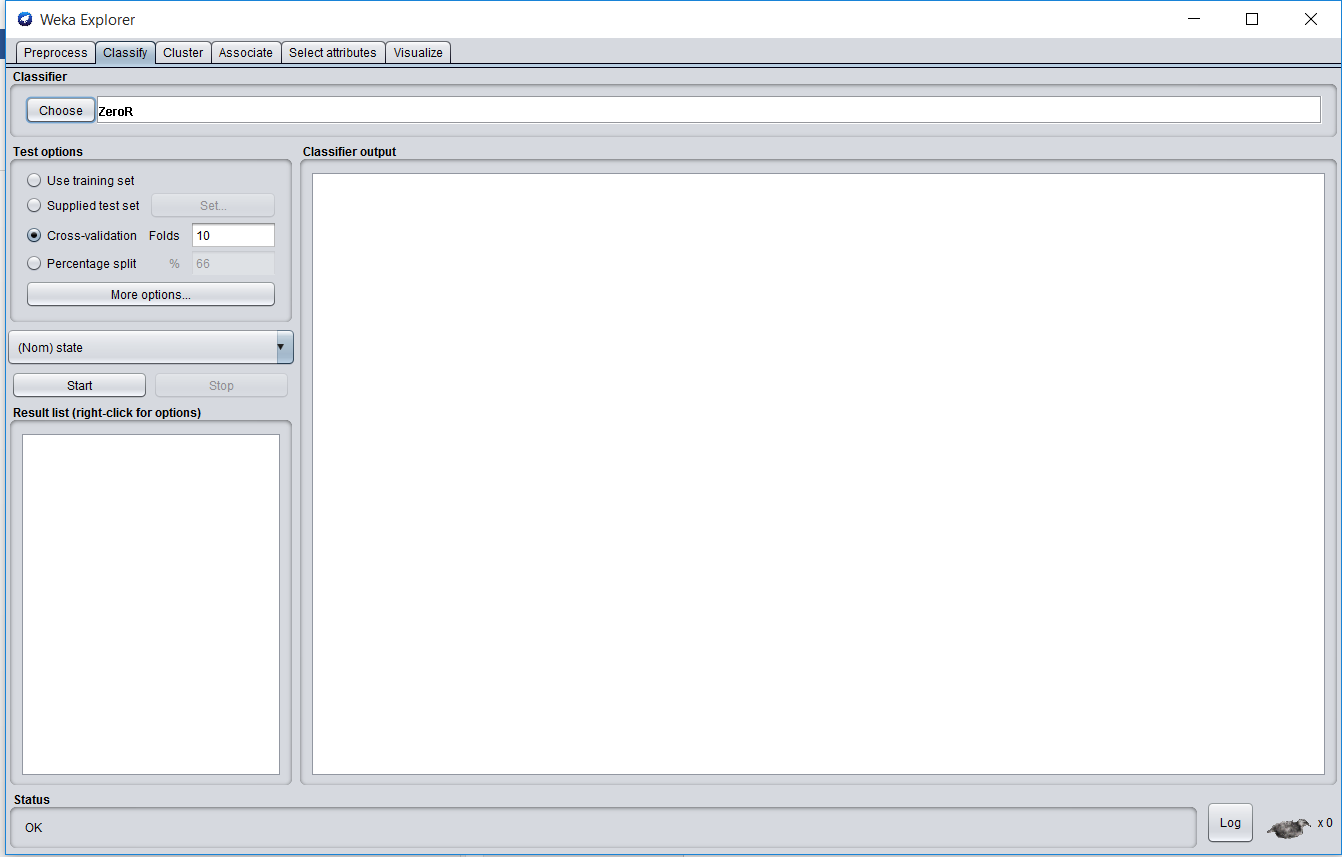
Với trạng thái mật độ giao thông là:

* + S1: có 13079 bản ghi dữ liệu
  + S2: có 15486 bản ghi dữ liệu
  + S3: 6392 bản ghi dữ liệu
  + S4: 9763 bản ghi dữ liệu
  + S5: 7018 bản ghi dữ liệu
  + S6: 1830 bản ghi dữ liệu

### Bước 4: Thực hiện phân lớp

Để thực hiện phân lớp dữ liệu với WEKA, từ màn hình WekaExplore chọn tab Classify.

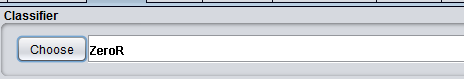
Ta có giao diện màn hình thực hiện phân lớp như sau:



Hình 9: Màn hình phân lớp dữ liệu

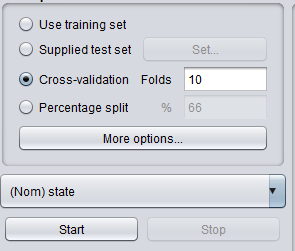
Trên màn hình thực hiện phân lớp sẽ có các lựa chọn sau:

* Chọn bộ dữ liệu huấn luyện



Hình 10: Chức năng chọn bộ phân lớp

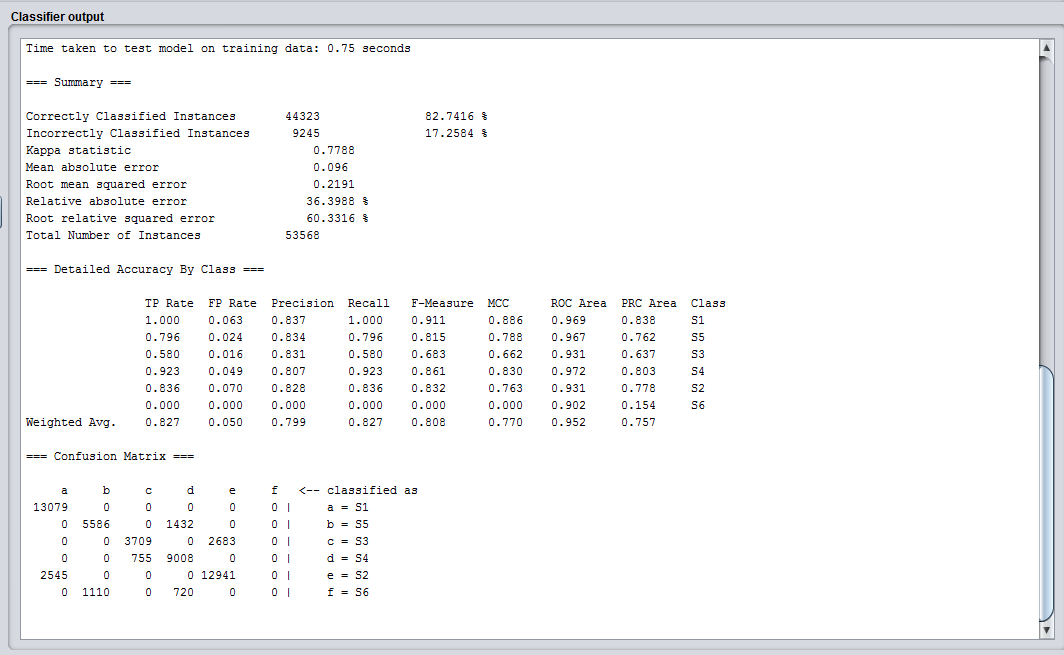
* Tại chức năng này, ta sẽ chọn bộ phân lớp muốn sử dụng. Tương ứng với các mô hình phần lớp, ta lựa chọn các bộ phân lớp trong WEKA như sau:
  + Mô hình phân lớp SVM: bộ phân lớp SMO
  + Mô hình phân lớp Navie Bayes: bộ phân lớp NavieBayes
  + Mô hình phân lớp với cây quyết định: bộ phân lớp J48
  + Mô hình phân lớp Neural Network: bộ phân lớp MultilayerPerceptron
* Lựa chọn phương thức test



Hình 11: Các phương thức test

Các phương thức test bao gồm:

* + Use training set: sử dụng bộ training làm dữ liệu test.
  + Supplied test set: sử dụng một bộ test riêng, bộ test này sẽ là 20% bộ dữ liệu được cắt ra từ dữ liệu thực nghiệm.
  + Cross-validation: chia bộ dữ liệu huấn luyện làm các phần nhỏ hơn, mỗi phần nhỏ sẽ được sử dụng để test, kết quả test sẽ là kết quả trung bình của các phần này.
  + Percentage split: cắt bộ dữ liệu huấn luyện theo một tỷ lệ % nhất định để làm bộ dữ liệu test.
* Kết quả test



Hình 12: Kết quả output phân lớp dữ liệu

## Kết quả thực nghiệm

### Cài đặt môi trường thực nghiệm

Thực nghiệm được tiến hành trên máy chủ Windows 10 có cấu hình được trình bày trong Bảng 4.2

|  |  |  |
| --- | --- | --- |
| **STT** | **Thông số phần cứng** | |
| 1 | CPU | Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz (4 CPUs), ~2.2GHz |
| 2 | RAM | 12Gb |
| 3 | SSD | 500Gb |
|  | **Thông số phần mềm** | |
| 4 | Hệ điều hành | Windows 10 |
| 5 | Công cụ | WEKA (Waikato Environment for Knowledge Analysis) |
| 6 | Gói hỗ trợ |  |

Bảng 3.1: Cấu hình máy chủ trong thực nghiệm

### Thực nghiệm test 70-30

#### Mô hình Decision Tree

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 81.2% | 100.0% | 89.6% | S1 |
| 79.9% | 80.0% | 79.9% | S2 |
| 81.3% | 52.5% | 63.8% | S3 |
| 78.3% | 96.3% | 86.4% | S4 |
| 80.2% | 63.8% | 71.1% | S5 |
| 79.7% | 82.6% | 81.1% | S6 |
| *Avg* | 75.3% | 80.0% | 77.6% |  |

Bảng 3.2: Test 70-30 - Kết quả test mô hình Decision Tree

#### Mô hình Super Vector Machines (SVM)

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 50.9% | 84.4% | 63.5% | S1 |
| 43.7% | 81.7% | 56.9% | S2 |
| 34.9% | 2.6% | 4.8% | S3 |
| 80.9% | 13.2% | 22.7% | S4 |
| 23.1% | 5.8% | 9.3% | S5 |
| 79.7% | 82.6% | 81.1% | S6 |
| *Avg* | 39.3% | 46.5% | 42.6% |  |

Bảng 3.3: Test 70-30 - Kết quả test mô hình SVM

#### Mô hình Naive Bayes

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 67.6% | 84.4% | 75.1% | S1 |
| 51.9% | 81.7% | 63.5% | S2 |
| 44.9% | 7.5% | 12.9% | S3 |
| 78.7% | 93.4% | 85.4% | S4 |
| 76.4% | 41.4% | 53.7% | S5 |
| 55.6% | 55.6% | 55.6% | S6 |
| *Avg* | 53.8% | 62.9% | 58.0% |  |

Bảng 3.4: Test 70-30 - Kết quả test mô hình Naïve Bayes

#### Mô hình Neural Network

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 80.3% | 83.8% | 82.0% | S1 |
| 72.0% | 80.6% | 76.1% | S2 |
| 80.3% | 82.5% | 81.4% | S3 |
| 79.4% | 94.9% | 86.5% | S4 |
| 81.4% | 64.1% | 71.7% | S5 |
| 79.1% | 85.7% | 82.3% | S6 |
| *Avg* | 76.4% | 77.2% | 76.8% |  |

Bảng 3.5: Test 70-30 - Kết quả test mô hình Neural Network

#### So sánh các kết quả chạy của SVM, Navies Bayes, J48 và Neural Network với số phân lớp là 6

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 75.3% | 39.3% | 53.8% | 76.4% |
| Recall | 80.0% | 46.5% | 62.9% | 77.2% |
| F-Measure | 77.6% | 42.6% | 58.0% | 76.8% |

Bảng 3.6: Test 70-30 - So sánh kết quả phân lớp sử dụng SVM, Navies Bayes, J48 và Neural Network với bộ phân lớp là 6.

### Thực nghiệm Test Cross validation với dữ liệu thông thường

Phương tức test Cross là phương thức test mà Bộ dữ liệu huấn luyện được chia làm nhiều phần khác nhau. Các phần này gọi là các Folds, mỗi một folds sẽ được sử dụng để làm 1 bộ test. Kết quả test Cross là kết quả test trung bình của các Folds:

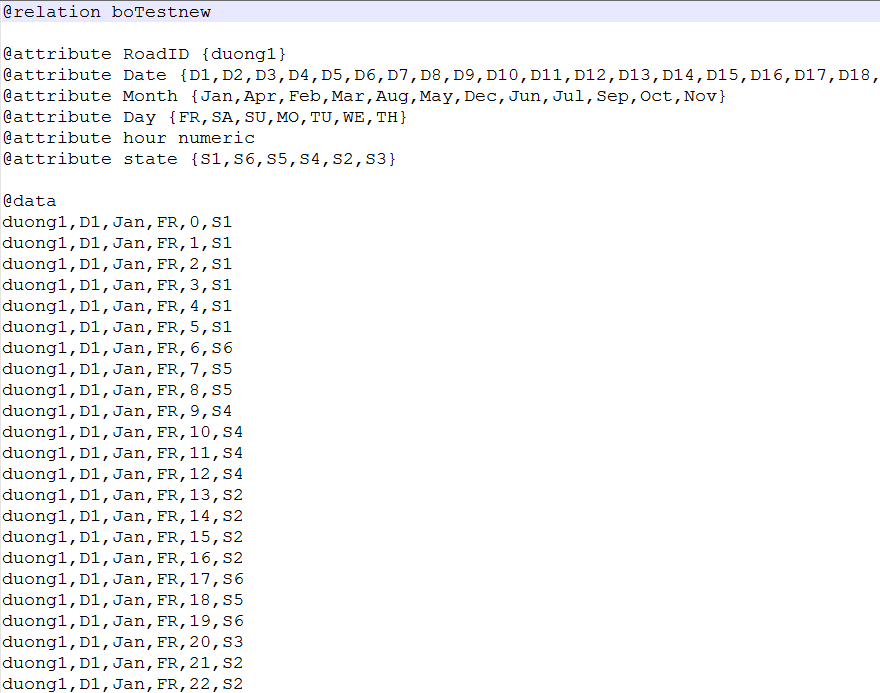
Dữ liệu thông thường trong bộ test này là dữ liệu không có thông tin ngày nghỉ lễ.

Ta có các tập dữ liệu tranning và test như sau:

Tool sử dụng: Weka (Waikato Environment for Knowledge Analysis)

Folds: 10

* Dữ liệu tranning:
  + Số lượng data: 17856
  + Số Attribute: 5
  + Số class: 6



Hình 17: Test Cross Validation – Dữ liệu training

#### Mô hình Decision Tree

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 81.0% | 100.0% | 89.5% | S1 |
| 81.1% | 97.9% | 88.7% | S2 |
| 81.3% | 52.5% | 63.8% | S3 |
| 80.6% | 89.6% | 84.9% | S4 |
| 66.1% | 18.9% | 29.4% | S5 |
| 79.4% | 100.0% | 88.5% | S6 |
| *Avg* | 78.6% | 80.0% | 79.3% |  |

Bảng 3.7: Test Cross Validation - Kết quả test mô hình Decision Tree

#### Mô hình Super Vector Machines (SVM)

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 52.3% | 85.8% | 65.0% | S1 |
| 44.7% | 82.5% | 58.0% | S2 |
| 34.9% | 2.6% | 4.8% | S3 |
| 81.0% | 11.3% | 19.8% | S4 |
| 22.6% | 7.5% | 11.3% | S5 |
| 24.1% | 29.5% | 26.5% | S6 |
| *Avg* | 40.1% | 47.2% | 43.4% |  |

Bảng 3.8: Test Cross Validation - Kết quả test mô hình SVM

#### Mô hình Naive Bayes

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 75.1% | 100.0% | 85.8% | S1 |
| 46.0% | 79.2% | 58.2% | S2 |
| 44.9% | 7.5% | 12.9% | S3 |
| 59.4% | 70.0% | 64.3% | S4 |
| 2.6% | 73.1% | 5.0% | S5 |
| 56.2% | 56.1% | 56.1% | S6 |
| *Avg* | 58.8% | 63.0% | 60.8% |  |

Bảng 3.9: Test Cross Validation - Kết quả test mô hình Naïve Bayes

#### Mô hình Neural Network

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 81.0% | 97.6% | 88.5% | S1 |
| 80.6% | 91.7% | 85.8% | S2 |
| 67.8% | 57.4% | 62.2% | S3 |
| 80.3% | 88.0% | 84.0% | S4 |
| 81.4% | 64.1% | 71.7% | S5 |
| 79.1% | 85.7% | 82.3% | S6 |
| *Avg* | 76.4% | 77.2% | 76.8% |  |

Bảng 3.10: Test Cross Validation - Kết quả test mô hình Neural Network

#### So sánh các kết quả chạy của SVM, Navies Bayes, J48 và Neural Network với số phân lớp là 6

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 78.6% | 40.1% | 58.8% | 76.4% |
| Recall | 80.0% | 47.2% | 63.0% | 77.2% |
| F-Measure | 79.3% | 43.4% | 60.8% | 76.8% |

Bảng 3.11: Test Cross - So sánh kết quả phân lớp sử dụng SVM, Navies Bayes, J48 và Neural Network.

### Thực nghiệm Test Cross validation với dữ liệu có xét đến ngày nghỉ lễ

Đẻ bám sát với thực tết hơn, ngoài thực nghiệm test với dữ liệu thông thường, ta còn thực hiện test với dữ liệu có thông tin các ngày nghỉ lễ đã được xây dựng ở chương trước.

Phương thức test được lựa chọn là Test Cross validation với bộ dữ liệu có chứa thông tin ngày nghỉ lễ:

#### Mô hình Decision Tree

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 79.4% | 100.0% | 88.5% | S1 |
| 79.4% | 77.2% | 78.3% | S2 |
| 41.3% | 52.5% | 46.2% | S3 |
| 4.1% | 78.8% | 7.8% | S4 |
| 79.8% | 79.8% | 79.8% | S5 |
| 79.4% | 100.0% | 88.5% | S6 |
| *Avg* | 74.3% | 79.4% | 76.8% |  |

Bảng 3.12: Test Cross Validation - Kết quả test mô hình Decision Tree

#### Mô hình Super Vector Machines (SVM)

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 51.2% | 75.0% | 60.9% | S1 |
| 32.1% | 76.5% | 45.2% | S2 |
| 32.1% | 12.3% | 17.8% | S3 |
| 71.2% | 0.3% | 0.6% | S4 |
| 22.6% | 7.5% | 11.3% | S5 |
| 23.2% | 29.5% | 26.0% | S6 |
| *Avg* | 39.9% | 40.2% | 40.0% |  |

Bảng 3.13: Test Cross Validation - Kết quả test mô hình SVM

#### Mô hình Naive Bayes

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 60.6% | 54.0% | 57.1% | S1 |
| 40.8% | 56.7% | 47.5% | S2 |
| 33.1% | 29.4% | 31.1% | S3 |
| 65.5% | 92.5% | 76.7% | S4 |
| 66.3% | 54.2% | 59.6% | S5 |
| 70.4% | 6.1% | 11.2% | S6 |
| *Avg* | 54.1% | 51.3% | 52.7% |  |

Bảng 3.14: Test Cross Validation - Kết quả test mô hình Naïve Bayes

#### Mô hình Neural Network

Kết quả test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chi tiết độ chính xác phân lớp | | | | |
|  | Precision | Recall | F-Measure | Class |
| 80.5% | 80.5% | 80.5% | S1 |
| 69.8% | 69.8% | 69.8% | S2 |
| 17.7% | 17.7% | 17.7% | S3 |
| 79.4% | 87.2% | 83.1% | S4 |
| 75.5% | 75.5% | 75.5% | S5 |
| 79.5% | 79.5% | 79.5% | S6 |
| *Avg* | 79.5% | 79.5% | 79.5% |  |

Bảng 3.15: Test Cross Validation - Kết quả test mô hình Neural Network

### Biểu đồ so sánh kết quả thực nghiệm giữa dữ liệu bình thường và dữ liệu có xét đến ngày nghỉ lễ

Biểu đồ 1: So sánh Dữ liệu thông thường và Dữ liệu có ngày nghỉ lễ

Dựa vào biểu đô ta thấy với

* Dữ liệu thông thường, độ chính xác cho kết quả cao hơn so với Dữ liệu có ngày nghỉ lễ.
* Độ chính xác của Neural network và Decission Tree đạt kết qura cao nhất, lên đến xấp chỉ 80%.
* Độ chính xác của SVM có kết quả thấp nhất, dưới 50%.
* Với giải thuật Navie Bayes và SVM thì độ chính xác khi Dữ liệu có ngày nghỉ lễ thấp hơn tương đối nhiều so với Dữ liệu thông thường trong khi Neural network và Decision Tree J48 độ chính xác là tương đồng nhau giữa hai bộ dữ liệu.

Như vậy có thể thấy việc dự đoán mật độ Giao thông với các bộ dữ liệu thì Neural Network và Decision Tree có tính chính xác cao hơn và phù hợp hơn với yêu cầu bài toán mà luận văn đưa ra.

### Thực nghiệm với các mức độ tắc đường khác nhau

#### Thực nghiệp với số lớp (mức độ tắc đường) là 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 80.5% | 58.9% | 57.5% | 75.0% |
| Recall | 80.5% | 71.9% | 69.7% | 76.1% |
| F-Measure | 80.5% | 64.8% | 63.0% | 75.5% |

#### Thực nghiệp với số lớp (mức độ tắc đường) là 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 80.1% | 31.0% | 48.7% | 71.6% |
| Recall | 80.2% | 51.4% | 56.2% | 71.6% |
| F-Measure | 80.1% | 38.7% | 52.2% | 71.6% |

#### Thực nghiệp với số lớp (mức độ tắc đường) là 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 82.3% | 38.9% | 49.9% | 76.5% |
| Recall | 81.7% | 53.5% | 58.6% | 76.1% |
| F-Measure | 82.0% | 45.0% | 53.9% | 76.3% |

#### Thực nghiệp với số lớp (mức độ tắc đường) là 6

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision  Tree | SVM | Navie Bayes | Neural Network |
| Precision | 78.6% | 40.1% | 54.8% | 76.4% |
| Recall | 80.0% | 47.2% | 51,3% | 77.2% |
| F-Measure | 79.3% | 43.4% | 53.0% | 76.8% |

### Biểu đồ so sánh kết quả thực nghiệm với các mật độ giao thông khác nhau

Biểu đồ 2: Kết quả thực nghiệm với các mật độ giao thông khác nhau

Dựa vào biểu đồ ta có thể thấy.

* Với bộ phân lớp Decision Tree J48, kết quả tương đối chính xác (giữ ở mức 80%) và không bị phụ thuộc vào số lượng class
* Với bộ phân lớp Neural Network, độ chính xác thấp hơn, nhưng giữ ổn định ở 76% và không bị ảnh hưởng nhiều vào số lượng class.
* Với bộ phân lớp SVM và Navie Bayes, độ chính xác thấp hơn và SVM thể hiện rõ sự yếu khi chạy với số lượng class lớn hơn.

Như vậy, bộ phân lớp nên được lựa chọn để sử dụng Dự đoán mật độ giao thông là Decision Tree J48.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TƯƠNG LAI

**Kết luận**

Dự đoán mật độ giao thông là một phần của Hệ thống giao thông thông minh, được xây dựng để phục vụ mục đích hỗ trợ người tham gia giao thông lựa chọn được hành trình phù hợp. Phương thức dự đoán mật độ giao thông đang được phát triển và tích hợp thêm các phương thức khác để có thể dự đoán chính xác hơn. Luận văn “Áp dụng mô hình phân lớp vào dự đoán mật độ giao thông” đã tiến hành khảo sát về mô hình phân lớp với Các bộ phân lớp khác nhau và áp dụng mô hình vào các dữ liệu thực nghiệm để đạt được các kết quả như mong muốn.

Kết quả đạt được trong luận văn:

* Giới thiệu được về Hệ thống giao thông thông minh và hệ thống con là Hệ thống dự đoán mật độ giao thông dựa trên mô hình Cây quyết định.
* Khảo sát và thống kê các mật độ độ giao thông trong khoảng thời gian nhất định.
* Nghiên cứu, tìm hiểu các hướng để tiếp cận mô hình phân lớp với cây quyết định.
* Chạy ra kết quả dự đoán mật độ giao thông cao với dữ liệu thực nghiệm, góp phần tăng độ quan trọng của Hệ thống dự đoán mật độ giao thông trong Hệ thống giao thông thông minh.

Những điều cần khắc phục:

* Việc áp dụng các sự cố (mưa, bão, …) vào dự đoán mật độ giao thông mới chỉ ở mức độ đơn giản, phần lớn chỉ là tăng hoặc giảm mật độ.

**Hướng phát triển tương lai**

Trong thời gian tới, luận văn sẽ tiếp tục nghiên cứu về viếc áp dụng mô hình phân lớp vào dự đoán mật độ giao thông. Đồng thời nghiên cứu kỹ hơn việc áp dụng các sự cố (mưa, bão,…) vào công tác dự đoán. Đi kèm việc nghiên cứu là bổ sung thêm các mô hình phân lớp khác để kết quả đầu ra đa dạng hơn và có sự so sánh nhất định giữa các mô hình.

# TÀI LIỆU THAM KHẢO

**Tiếng Anh**

1. *Naive Bayes Classifiers and Document Classification*- Brandon Malone . January 24, 2014.
2. *Decision Tree Analysis on J48 Algorithm for Data Mining*- Dr. Neeraj Bhargava, Girja Sharma, Dr. Ritu Bhargava, Manish Mathuria - Volume 3, Issue 6, June.
3. *Support Vector Machine (and Statistical Learning Theory) Tutorial Jason Weston  
   NEC Labs America 4 Independence Way*, Princeton, USA.  
   [jasonw@nec-labs.com](mailto:jasonw@nec-labs.com).
4. *Artifical Neural Networks* - Ani1 K. Jain Michigan State University Jianchang M a o K.M. Mohiuddin ZBMAZmadenResearch Center
5. *Explaining International IT Application Leaderhip: Intelligent Transportation Systems* - Stephen Ezell. January 2010.
6. *A Tutorial on Bayesian classifier A Tutorial on Bayesian classifier with W*EKA - MING-CHANG LEE - Department of Information Management Yu Da College of Business. March 28, 2006.
7. *Induction of Decision Trees* - J.R. QUINLAN - Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia.
8. *Traffic simulation with consideration of driver models, theory and examples* - J. Ludmann, D. Neunzig, and M. Weilkes - Veh. Syst. Dyn. 27, 491-516 (1997).
9. *The benefits of intelligent transport systems: modelling the effects of different its systems -* Vanderschuren, M.J.W.A. – Jully ,2003
10. *Behavioral Policies and Teen Traffic Safety* , Thomas S. Dee and William N. Evans – May, 2001