**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

**BÁO CÁO ĐỒ ÁN MÔN HỌC PROJECT 3**

## Tên đề tài: Ứng dụng phương pháp SVR để dự đoán chi phí và thời gian hoàn thành dự án phần mềm

|  |  |
| --- | --- |
| Giảng viên hướng dẫn | PGS. TS. Huỳnh Quyết Thắng |
| Sinh viên thực hiện | Nguyễn Đăng Trung Kiên 20173203  Nguyễn Văn Trung 20173424 |
|  |  |

**HÀ NỘI 2020**

**HÀ NỘI 2020**

**Hà Nội, 06/2009**

MỤC LỤC

[Tên đề tài: Ứng dụng phương pháp SVR để dự đoán chi phí và thời gian hoàn thành dự án phần mềm 1](file:///E:\download2\Baocaoproject3_SVR.docx#_Toc62485123)

[Giảng viên hướng dẫn 1](file:///E:\download2\Baocaoproject3_SVR.docx#_Toc62485124)

[PGS. TS. Huỳnh Quyết Thắng 1](file:///E:\download2\Baocaoproject3_SVR.docx#_Toc62485125)

[Sinh viên thực hiện 1](file:///E:\download2\Baocaoproject3_SVR.docx#_Toc62485126)

[DANH MỤC BẢNG VÀ HÌNH VẼ 4](#_Toc62485127)

[DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT 5](#_Toc62485128)

[MỞ ĐẦU 6](#_Toc62485129)

[1. Thực trạng 6](#_Toc62485130)

[2. Định hướng giải pháp 6](#_Toc62485131)

[CƠ SỞ LÝ THUYẾT 7](#_Toc62485132)

[1.1. Tổng quan về quản lý dự án phần mềm 7](#_Toc62485133)

[1.1.1. Khái niệm và lịch sử phát triển 7](#_Toc62485134)

[1.1.2. Quy trình quản lý dự án trong phần mềm 7](#_Toc62485135)

[1.1.3. Các hoạt động chính trong quản lý dự án phần mềm 8](#_Toc62485136)

[1.2. Phương pháp truyền thống Quản trị giá trị thu được (EVM) 10](#_Toc62485137)

[1.2.1. Giới thiệu về EVM 10](#_Toc62485138)

[1.2.2. Các độ đo trong quản trị giá trị thu được: 11](#_Toc62485139)

[1.2.3. Sử dụng kỹ thuật quản trị giá trị thu được trong đánh giá tiến độ hiện tại 12](#_Toc62485140)

[1.2.4. Sử dụng kỹ thuật quản trị giá trị thu được trong việc dự đoán thời gian hoàn thành và chi phí dự án 13](#_Toc62485141)

[1.2.5. Ý nghĩa của EVM 14](#_Toc62485142)

[1.3. Phương pháp SVM 14](#_Toc62485143)

[1.3.1. Support Vector Machine (SVM) 14](#_Toc62485144)

[1.3.2. Kernel trong SVM – RBF Kernel 17](#_Toc62485145)

[1.3.3. Support Vector Regression (SVR) 19](#_Toc62485146)

[ÁP DỤNG THUẬT TOÁN SVR ĐỂ ƯỚC LƯỢNG THỜI GIAN VÀ CHI PHÍ HOÀN THÀNH DỰ ÁN PHẦN MỀM 21](#_Toc62485147)

[1.4. Bài toán đánh giá mức độ hoàn thành dự án phần mềm 21](#_Toc62485148)

[1.4.1. Dự đoán thời gian hoàn thành dự án 21](#_Toc62485149)

[1.4.2. Dự đoán chi phí hoàn thành dự án 21](#_Toc62485150)

[1.5. Xây dựng tập dữ liệu 22](#_Toc62485151)

[1.5.1. Tập dữ liệu kế thừa từ các khóa trước 22](#_Toc62485152)

[1.5.2. Thu thập thêm dữ liệu 23](#_Toc62485153)

[1.6. Xây dựng mô hình SVR 23](#_Toc62485154)

[1.6.1. Dữ liệu huấn luyện, dữ liệu test 23](#_Toc62485155)

[1.6.2. Lựa chọn tham số C, Gamma 24](#_Toc62485156)

[1.6.3. Đánh giá kết quả dự đoán 24](#_Toc62485157)

[CÀI ĐẶT VÀ ĐÁNH GIÁ THỰC NGHIỆM 25](#_Toc62485158)

[1.7. Môi trường thực nghiệm 25](#_Toc62485159)

[1.8. Kết quả thực nghiệm 25](#_Toc62485160)

[1.9. Đánh giá mô hình 27](#_Toc62485161)

[1.10. So sánh chương trình của K61 28](#_Toc62485162)

[1.11. So sánh với phương pháp EVM 29](#_Toc62485163)

[KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN 30](#_Toc62485164)

[1.12. Kết luận 30](#_Toc62485165)

[1.13. Định hướng phát triển 30](#_Toc62485166)

[DANH MỤC TÀI LIỆU THAM KHẢO 31](#_Toc62485167)

[PHỤ LỤC 32](#_Toc62485168)

# DANH MỤC BẢNG VÀ HÌNH VẼ

**Danh mục hình vẽ**

[Hình 1: Sử dụng EVM dự đoán thời gian và chi phí hoàn thành dự án 13](#_Toc60225914)

[Hình 2: Classify SVM 14](https://husteduvn-my.sharepoint.com/personal/kien_ndt173203_sis_hust_edu_vn/Documents/Baocaoproject3moi.docx#_Toc60225915)

[Hình 3: Margin SVM 15](https://husteduvn-my.sharepoint.com/personal/kien_ndt173203_sis_hust_edu_vn/Documents/Baocaoproject3moi.docx#_Toc60225916)

[Hình 4: Tham số C trong SVM 17](https://husteduvn-my.sharepoint.com/personal/kien_ndt173203_sis_hust_edu_vn/Documents/Baocaoproject3moi.docx#_Toc60225917)

[Hình 5: Kernel 18](https://husteduvn-my.sharepoint.com/personal/kien_ndt173203_sis_hust_edu_vn/Documents/Baocaoproject3moi.docx#_Toc60225918)

[Hình 6: Tham số gamma trong RBF Kernel 19](#_Toc60225919)

[Hình 7: SVR 20](#_Toc60225920)

[Hình 8: Dữ liệu sau xử lí 22](#_Toc60225921)

[Hình 9: Dữ liệu input của chương trình 23](https://husteduvn-my.sharepoint.com/personal/kien_ndt173203_sis_hust_edu_vn/Documents/Baocaoproject3moi.docx#_Toc60225922)

[Hình 10: x\_train 24](#_Toc60225923)

[Hình 11: y\_train 24](#_Toc60225924)

[Hình 12: Kết quả dự đoán thời gian hoàn thành dự án tại các mốc thời gian K62 25](#_Toc60225925)

[Hình 13: Biểu đồ kết quả dự đoán và kết quả thực tế thời gian hoàn thành dự án 26](#_Toc60225926)

[Hình 14: Kết quả dự đoán chi phí hoàn thành dự án tại các mốc thời gian K62 26](#_Toc60225927)

[Hình 15: Biểu đồ kết quả dự đoán và kết quả thực tế chi phí hoàn thành dự án 27](#_Toc60225928)

[Hình 16: Kết quả dự đoán thời gian hoàn thành dự án của K61 28](#_Toc60225929)

[Hình 17: Kết quả dự đoán chi phí hoàn thành dự án của K61 29](#_Toc60225930)

# DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **Ký hiệu, viết tắt** | **Ý nghĩa tiếng Anh** | **Ý nghĩa tiếng Việt** |
| AC | Actual Cost | Chi phí thực tế |
| AT | Actual Time | Thời gian thực tế |
| EV | Earned Value | Giá trị thu được |
| EVM | Earned Value Management | Quản trị giá trị thu được |
| PV | Planned Value | Giá trị dự kiến |
| TOTAL\_AT | Total Time | Tổng thời gian thực hiện dự án |
| TOTAL\_EV | Total Earned Value | Tổng giá trị thu được |
| TOTAL\_AC | Total Cost | Tổng chi phí |
| XT | Current Progress Percentage | Phần trăm tiến độ |

# MỞ ĐẦU

## 1. Thực trạng

Trong giai đoạn hiện nay, toàn thế giới chạy đua trong cuộc Cách mạng Công nghiệp 4.0. Các công ty buộc phải chạy đua công nghệ nếu không muốn bị tụt hậu. Số lượng các dự án phần mềm bùng nổ mạnh mẽ, đi cùng với chi phí thực hiện khổng lồ. Mọi sai sót khi thực hiện có thể dẫn đến thiệt hại nặng nề.

Không phải dự án nào cũng thành công, rất nhiều trong số đó có kết quả không như kỳ vọng. Theo thống kê, 35% số dự án phần mềm thất bại vì các lý do: thời hạn, chi phí, chất lượng (không đáp ứng được nghiệp vụ, khó sử dụng, không tin cậy…). Có thể nhận thấy, dự án thất bại chủ yếu do quản trị không tốt. Ngay từ khi thiết lập dự án, không có kế hoạch tốt ví dụ như không xác định rõ mục tiêu, không thống nhất rõ yêu cầu, thời hạn và kinh phí không thực tế. Trong khi thực hiện dự án, việc phân tích và thiết kế sai lầm, không được văn bản hoá, điều phối nguồn lực không đúng, thiếu công cụ hỗ trợ.

Nếu như dự đoán được thời gian và chi phí hoàn thành dự án, người quản lý dự án và nhóm dự án quản lý dự án một cách chủ động, tính trước được các rủi ro và nguy cơ. Vì vậy, việc tìm ra các phương pháp giúp dự đoán chi phí, thời gian của dự án ở các thời điểm là điều vô cùng quan trọng.

## 2. Định hướng giải pháp

Quản trị giá trị thu được (EVM) là một trong những kỹ thuật nổi tiếng để kiểm soát thời gian và chi phí của một dự án. Phương pháp này đánh giá theo dõi dự án trên dựa trên một tập hợp các độ đo nhằm đưa ra các dự đoán sớm cho người quản trị dự án. Tuy nhiên, phương pháp này cũng có hạn chế là dự đoán thiếu tính tin cậy trong giai đoạn sớm của dự án. Vì vậy cần có giải pháp tốt hơn. Một trong số các cách hiệu quả đó là kết hợp EVM cùng các phương pháp học máy

Trong phạm vi của môn học Project 3, kế thừa công trình nghiên cứu của K61, chúng em chọn đề tài “**Ứng dụng phương pháp SVR và EVM để dự đoán chi phí và thời gian hoàn thành dự án phần mềm**”. Mục đích tiếp tục tối ưu hơn thuật toán, và tạo ra mô hình chính xác hơn.

# CƠ SỞ LÝ THUYẾT

## Tổng quan về quản lý dự án phần mềm

### Khái niệm và lịch sử phát triển

Quản lý dự án phần mềm việc lên kế hoạch có tính [khoa học](https://vi.wikipedia.org/wiki/Khoa_h%E1%BB%8Dc)  trong quá trình quản lý các dự án [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm). Nó chính là quy trình quản lý dự án theo các chính sách được lên kế hoạch, theo dõi và kiểm soát.

Các [công ty](https://vi.wikipedia.org/wiki/C%C3%B4ng_ty" \o "Công ty) nhanh chóng nhận ra việc sử dụng lập trình [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) dễ hơn so với [phần cứng](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_c%E1%BB%A9ng" \o "Phần cứng) vì vậy ngành công nghiệp sản xuất [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) nhanh chóng phát triển trong giai đoạn từ thập niên 1970 đến thập niên 1980. Để quản lý các nỗ lực phát triển mới, các [công ty](https://vi.wikipedia.org/wiki/C%C3%B4ng_ty" \o "Công ty) ứng dụng các phương thức quản lý [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm), nhưng quá trình thử nghiệm bị chậm theo thực thi, đặc biệt là sự mẫu thuẫn xảy ra trong "vùng xám" giữa các đặc tả người dùng và [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) được chuyển giao. Để tránh các vấn đề trên, các phương thức quản lý dự án [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) tập trung vào các yêu cầu người dùng trong các sản phẩm phần mềm theo mô hình nổi tiếng đó là [mô hình thác nước](https://vi.wikipedia.org/wiki/M%C3%B4_h%C3%ACnh_th%C3%A1c_n%C6%B0%E1%BB%9Bc" \o "Mô hình thác nước). Theo tổ chức [IEEE](https://vi.wikipedia.org/wiki/IEEE), một số nguyên nhân dẫn đến sự thất bại trong việc quản lý dự án như sau:

1. Dự án không có tính thực tế và không khớp
2. Ước tính không chính xác nguồn lực cần thiết cho dự án
3. Xác định yêu cầu hệ thống không đúng
4. Báo cáo tình trạng dự án sơ sài
5. Không quản lý độ rủi ro
6. Việc giao tiếp [khách hàng](https://vi.wikipedia.org/wiki/Kh%C3%A1ch_h%C3%A0ng" \o "Khách hàng), [người sử dụng](https://vi.wikipedia.org/w/index.php?title=Ng%C6%B0%E1%BB%9Di_s%E1%BB%AD_d%E1%BB%A5ng&action=edit&redlink=1" \o "Người sử dụng (trang chưa được viết)) và người phát triển [dự án](https://vi.wikipedia.org/wiki/D%E1%BB%B1_%C3%A1n" \o "Dự án) không tốt
7. Sử dụng công nghệ chưa phát triển
8. Không có khả năng xử lý độ phức tạp của dự án
9. Phát triển thực hành không có hệ thống
10. Thiếu kinh nghiệm trong việc quản lý dự án

### Quy trình quản lý dự án trong phần mềm

Quy trình quản lý dự án phần mềm là quy trình vận dụng những kiến thức, kỹ năng và kỹ thuật công nghệ vào hoạt động của dự án để đạt được mục tiêu của dự án đặt ra. Những ứng dụng này được đưa vào [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) theo một tiêu chuẩn hóa của quản lý dự án theo tiêu chuẩn [PMI](https://vi.wikipedia.org/wiki/Vi%E1%BB%87n_Qu%E1%BA%A3n_l%C3%BD_D%E1%BB%B1_%C3%A1n).

Để đảm bảo dự án thành công, các thành viên dự án phải đảm bảo:

* Lựa chọn quy trình phù hợp để đạt được mục tiêu của [dự án](https://vi.wikipedia.org/wiki/D%E1%BB%B1_%C3%A1n" \o "Dự án)
* Tuân theo các yêu cầu để đáp ứng được nhu cầu và mong đợi của các bên liên quan.
* Cân bằng được các yêu cầu (nhân tố) cạnh tranh trong dự án như: phạm vi công việc, [ngân sách](https://vi.wikipedia.org/wiki/Ng%C3%A2n_s%C3%A1ch" \o "Ngân sách), tiến độ, chất lượng, rủi ro, thay đổi. Tùy theo quy mô của từng dự án mà các mỗi giai đoạn lại có thể gồm những quy trình nhỏ hơn.

Ngoài các lợi ích chiến lược nêu trên [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) còn cung cấp đầy đủ các tính năng hệ thống. Việc bảo mật được tiến hành một cách tuyệt đối nghiêm ngặt. Việc phân quyền được cụ thể đến từng vai trò của người sử dụng.

Quy trình kiểm tra và giám sát dự án quản lý phần mềm bao gồm 5 giai đoạn.

1. Khởi tạo dự án (Initiating): Giai đoạn này thực hiện việc định nghĩa một dự án mới hoặc một phát sinh (hoặc trộn lẫn) mới của một [dự án](https://vi.wikipedia.org/wiki/D%E1%BB%B1_%C3%A1n" \o "Dự án) có sẵn như: Xác định yêu cầu của dự án, mức độ ưu tiên của dự án, phân tích các yêu cầu đầu tư, phân công trách nhiệm cho các bộ phận triển khai.

2. Lập kế hoạch dự án (Planning): Giai đoạn này yêu cầu thiết lập phạm vi công viêc của dự án, điều chỉnh lại mục tiêu và xác định đường đi tới mục tiêu đó.

3. Triển khai (Executing): Giai đoạn này thực hiện hoàn thành các công việc được xác định trong phần lập kế hoạch để đảm bảo các yêu cầu của dự án.

4. Giám sát và kiểm soát (Monitoring & Control): Giai đoạn này yêu cầu việc theo dõi, rà soát và điều chỉnh lại tiến độ và khả năng thực hiện của dự án. Theo dõi các rủi ro, thay đổi, phát sinh trong quá trình thực hiện và có những đề xuất điều chỉnh kịp thời.

5. Kết thúc (Closing): Giai đoạn này thực hiện để kết thúc tất cả các hoạt động của dự án để chính thức đóng lại dự án.

### Các hoạt động chính trong quản lý dự án phần mềm

Xác định các bước thực hiện dự án phần mềm

**Xác định yêu cầu chung**

Trước tiên, cần xác định các [yêu cầu chức năng](https://vi.wikipedia.org/w/index.php?title=Y%C3%AAu_c%E1%BA%A7u_ch%E1%BB%A9c_n%C4%83ng&action=edit&redlink=1" \o "Yêu cầu chức năng (trang chưa được viết)) (công việc phần mềm thực hiện) cũng như phi chức năng (công nghệ dùng để phát triển phần mềm, sử dụng trong [hệ điều hành](https://vi.wikipedia.org/wiki/H%E1%BB%87_%C4%91i%E1%BB%81u_h%C3%A0nh" \o "Hệ điều hành)) của [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm). Tiếp theo cần xác định rõ [tài nguyên](https://vi.wikipedia.org/wiki/T%C3%A0i_nguy%C3%AAn" \o "Tài nguyên) cần thiết để xây dựng [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm). [Tài nguyên](https://vi.wikipedia.org/wiki/T%C3%A0i_nguy%C3%AAn" \o "Tài nguyên) ở đây có thể gồm có nhân tố con người, các thành phần, phần mềm có thể sử dụng lại, các phần cứng hoặc công cụ có sẵn cần dùng đến; trong đó nhân tố con người là quan trọng nhất. Điều cuối cùng là xác định thời gian cần thiết để thực hiện dự án. Trong quá trình này cần phải nắm bắt được bài toán thực tế cần giải quyết cũng như các hoạt động mang tính nghiệp vụ của khách hàng để có thể xác định rõ ràng yêu cầu chung của đề án, xem xét dự án có khả thi hay không.

**Viết đề án**

Viết đề án là quá trình xây dựng tài liệu mô tả đề án để xác định phạm vi của dự án, trách nhiệm của những người tham gia dự án; là cam kết giữa người quản lý dự án, người tài trợ dự án và khách hàng. Nội dung của tài liệu mô tả đề án thường có những nội dung sau:

* Bối cảnh thực hiện dự án
* Mục đích và mục tiêu của dự án
* Phạm vi dự án
* Nguồn nhân lực tham gia dự án
* Ràng buộc thời gian thực hiện dự án
* Ràng buộc kinh phí
* Ràng buộc công nghệ phát triển
* [Chữ ký](https://vi.wikipedia.org/wiki/Ch%E1%BB%AF_k%C3%BD) các bên liên quan tới dự án

**Lập kế hoạch thực hiện dự án**

Lập kế hoạch thực hiện dự án là hoạt động diễn ra trong suốt quá trình từ khi bắt đầu thực hiện dự án đến khi bàn giao [sản phẩm](https://vi.wikipedia.org/wiki/S%E1%BA%A3n_ph%E1%BA%A9m" \o "Sản phẩm) với nhiều loại kế hoạch khác nhau nhằm hỗ trợ kế hoạch chính của dự án [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm) về lịch trình và [ngân sách](https://vi.wikipedia.org/wiki/Ng%C3%A2n_s%C3%A1ch" \o "Ngân sách).

**Các loại kế hoạch thực hiện dự án**

* Kế hoạch đảm bảo chất lượng: Mô tả các chuẩn, các quy trình được sử dụng trong dự án.
* Kế hoạch thẩm định: Mô tả các phương pháp, nguồn lực, lịch trình thẩm định hệ thống.
* Kế hoạch quản lý cấu hình: Mô tả các thủ tục, cấu trúc quản lý cấu hình được sử dụng.
* Kế hoạch bảo trì: Dự tính các yêu cầu về hệ thống, chi phí, nỗ lực cần thiết cho bảo trì.
* Kế hoạch phát triển đội ngũ: Mô tả kĩ năng và kinh nghiệm của các thành viên trong nhóm dự án sẽ phát triển như thế nào.

**Quy trình lập kế hoạch thực hiện dự án**

* Thiết lập các ràng buộc của dự án: thời gian, nhân lực, ngân sách
* Đánh giá bước đầu về các "tham số" của dự án: quy mô, độ phức tạp, nguồn lực
* Xác định các mốc thời gian trong thực hiện dự án và sản phẩm thu được ứng với mỗi mốc thời gian
* Trong khi dự án chưa hoàn thành hoặc chưa bị hủy bỏ thì thực hiện lặp đi lặp lại các công việc sau:

1. Lập lịch thực hiện dự án
2. Thực hiện các hoạt động theo lịch trình
3. Theo dõi sự tiến triển của dự án, so sánh với lịch trình
4. Đánh giá lại các tham số của dự án
5. Lập lại lịch thực hiện dự án cho các tham số mới
6. Thỏa thuận lại các ràng buộc và sản phẩm bàn giao của mỗi mốc thời gian
7. Nếu có vấn đề nảy sinh thì xem xét lại các kĩ thuật khởi đầu đưa ra các biện pháp cần thiết

**Cấu trúc kế hoạch thực hiện dự án**

* Tổ chức dự án
* Phân tích các rủi ro
* Yêu cầu về tài nguyên [phần cứng](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_c%E1%BB%A9ng" \o "Phần cứng), [phần mềm](https://vi.wikipedia.org/wiki/Ph%E1%BA%A7n_m%E1%BB%81m" \o "Phần mềm)
* [Phân công](https://vi.wikipedia.org/w/index.php?title=Ph%C3%A2n_c%C3%B4ng&action=edit&redlink=1) công việc
* [Lập lịch](https://vi.wikipedia.org/w/index.php?title=L%E1%BA%ADp_l%E1%BB%8Bch&action=edit&redlink=1) dự án
* Cơ chế [kiểm soát](https://vi.wikipedia.org/w/index.php?title=Ki%E1%BB%83m_so%C3%A1t&action=edit&redlink=1" \o "Kiểm soát (trang chưa được viết)) và [báo cáo](https://vi.wikipedia.org/wiki/B%C3%A1o_c%C3%A1o" \o "Báo cáo).

## Phương pháp truyền thống Quản trị giá trị thu được (EVM)

### Giới thiệu về EVM

Từ bắt đầu một dự án và trong suốt các giai đoạn của nó, người quản lý dự án và đội ngũ quản lý dự án phải giải quyết nhiều câu hỏi. Trong đó, điều mà cần quan tâm nhất là đưa ra các dự đoán của dự án trong tương lai, từ đó, có thể đưa ra các điều chỉnh phù hợp ở hiện tại. Có rất nhiều kỹ thuật giúp đưa ra các dự đoán về dự án, trong đó kỹ thuật “Quản trị giá trị thu được”, hay EVM (Earned Value Method), là kỹ thuật rộng rãi và mang lại lợi ích đáng kể cho người quản lý dự án và đội quản lý dự án.

Những điểm đặc trưng chủ yếu của việc triển khai thực hiện thuật quản lý giá trị thu được bao gồm:

1. Một bản [kế hoạch dự án](https://vi.wikipedia.org/w/index.php?title=K%E1%BA%BF_ho%E1%BA%A1ch_d%E1%BB%B1_%C3%A1n&action=edit&redlink=1" \o "Kế hoạch dự án (trang chưa được viết)) (được lập trước khi khởi công) xác định công việc phải hoàn thành, trong đó bao gồm cả kế hoạch về chi phí thực hiện (tức là bản [dự toán](https://vi.wikipedia.org/w/index.php?title=D%E1%BB%B1_to%C3%A1n&action=edit&redlink=1" \o "Dự toán (trang chưa được viết))) và kế hoạch về thời gian thực hiện (tức là bản [tiến độ](https://vi.wikipedia.org/w/index.php?title=Ti%E1%BA%BFn_%C4%91%E1%BB%99&action=edit&redlink=1" \o "Tiến độ (trang chưa được viết))).
2. Giá trị kinh phí dự kiến (cấp theo kế hoạch dự án) cho một công việc tại thời điểm kiểm soát dự án (còn gọi là thời điểm báo cáo), được gọi là [giá trị dự kiến](https://vi.wikipedia.org/w/index.php?title=Gi%C3%A1_tr%E1%BB%8B_d%E1%BB%B1_ki%E1%BA%BFn&action=edit&redlink=1" \o "Giá trị dự kiến (trang chưa được viết)) PV (Planned Value) hoặc [Dự toán ngân quỹ chi phí cho công việc theo tiến độ](https://vi.wikipedia.org/w/index.php?title=D%E1%BB%B1_to%C3%A1n_ng%C3%A2n_qu%E1%BB%B9_chi_ph%C3%AD_cho_c%C3%B4ng_vi%E1%BB%87c_theo_ti%E1%BA%BFn_%C4%91%E1%BB%99&action=edit&redlink=1" \o "Dự toán ngân quỹ chi phí cho công việc theo tiến độ (trang chưa được viết)) BCWS (tiếng Anh là Budgeted Cost of Work Scheduled),
3. Giá trị kinh phí (tức chi phí) theo dự toán (tức là kế hoạch trước khởi công) của phần khối lượng công việc thực tế đã hoàn thành tính tới thời điểm báo cáo (của [công việc](https://vi.wikipedia.org/wiki/C%C3%B4ng_vi%E1%BB%87c) được theo dõi), gọi là giá trị thu được EV (Earned Value) hoặc [Dự toán ngân quỹ chi phí cho công việc đã thực hiện](https://vi.wikipedia.org/w/index.php?title=D%E1%BB%B1_to%C3%A1n_ng%C3%A2n_qu%E1%BB%B9_chi_ph%C3%AD_cho_c%C3%B4ng_vi%E1%BB%87c_%C4%91%C3%A3_th%E1%BB%B1c_hi%E1%BB%87n&action=edit&redlink=1" \o "Dự toán ngân quỹ chi phí cho công việc đã thực hiện (trang chưa được viết)) BCWP (tiếng Anh là Budgeted Cost of Work Performed).

Quản lý giá trị thu được, được triển khai thực hiện ở các dự án lớn hoặc phức tạp, còn bao gồm nhiều tính năng hơn, chẳng hạn như các chỉ số và dự báo về chi phí thực hiện CPI (vượt [ngân sách](https://vi.wikipedia.org/wiki/Ng%C3%A2n_s%C3%A1ch) (tức là vượt ngân quỹ, đồng nghĩa bị lỗ) hoặc (dưới ngân quỹ, có lãi) trong vòng [ngân quỹ](https://vi.wikipedia.org/wiki/Ng%C3%A2n_s%C3%A1ch)) và tiến độ thực hiện SPI (chậm [tiến độ](https://vi.wikipedia.org/w/index.php?title=Ti%E1%BA%BFn_%C4%91%E1%BB%99&action=edit&redlink=1" \o "Tiến độ (trang chưa được viết)) hoặc vượt tiến độ (tức là trước thời hạn)). Tuy nhiên, yêu cầu cơ bản nhất của một hệ thống Quản lý giá trị thu được là nó định lượng được tiến trình thực hiện dự án bằng cách sử dụng giá trị dự kiến PV và giá trị thu được EV.

### Các độ đo trong quản trị giá trị thu được:

EVM tích hợp ba nhân tố chính của quản lý dự án: chi phí (cost), lịch trình tiến độ (schedule) và phạm vi (scope) vào trong phương pháp tính của mình. Chúng ta có các thành phần chính như sau:

* **PD (Planned Duration):** thời gian dự kiến dự án sẽ hoàn thành.
* **BAC (Budget at Completion):** tổng chi phí của dự án.
* **AT (Actual Time):** mốc thời gian thực tế của dự án.
* **PV (Planned Value):** giá trị dự kiến thu được tại mốc thời gian **AT**.
* **EV (Earned Value):** giá trị thu được thực tế tại mốc thời gian **AT.**
* **AC (Actual Cost):** chi phí thực tế bỏ ra tại mốc thời gian **AT.**
* **ES (Earned Schedule):** thời điểm mà giá trị **ES** dự kiến thu về được.
* **RD (Real Duration):** thời gian thực tế của dự án.
* **RC (Real Cost):** chi phí thực tế của dự án.
* **EAC(t) (Estimated time At Completion):** thời gian dự kiến tại mốc thời gian **AT.**
* **EAC (Estimated cost At Completion):** chi phí dự kiến tại mốc thời gian **AT.**
* **SV (Schedule Variance) = EV – PV.**
* **SPI (Schedule Performance Index) = EV / PV.**
* **SV(t) (Schedule Variance (time)) = ES – AT.**
* **SPI(t) (Schedule Performance Index (time)) = ES / AT.**
* **CV (Cost Variance) = EV – AC.**
* **CPI (Cost Performance Index) = EV / AC.**
* **SCI (Schedule Cost Index) = SPI \* CPI.**
* **SCI(t) (Schedule Cost Index (time)) = SPI(t) \* CPI.**

Các thành phần trên được dùng để đo đạc tiến độ và chi phí của dự án tại một thời điểm bất kỳ. Dựa vào các số liệu thu được và tính toán được, chúng ta sẽ áp dụng các công thức để có thể đưa ra các giá trị dự đoán phù hợp với tình hình dự án tại thời điểm bất kỳ. Trong các thành phần trên thì **PV, AC** và **EV** là ba thành phần chính tương ứng với ba nhân tố đã nói ở trên.

Từ các nhân tố trên, chúng ta sẽ tính toán được các đường hiệu suất cơ sở (Performance Baseline) để có thể đánh giá tình hình dự án hiện tại, qua đó sẽ đưa ra các quyết định tiếp theo cho dự án

### Sử dụng kỹ thuật quản trị giá trị thu được trong đánh giá tiến độ hiện tại

Giá trị thu được là tổng của các giá trị dự kiến đã được thực hiện xong (hoàn thành), từ khi khởi công dự án đến thời điểm hiện hành (thời điểm theo dõi dự án).{\displaystyle {\begin{aligned}\mathrm {EV} &=\sum \_{\mathrm {KhoiCong} }^{\mathrm {HienHanh} }\mathrm {PV(DaHoanThanh)} \\\end{aligned}}}

Giá trị thu được thực chất là phần chi phí đã được bỏ ra thực sự còn tồn lại trong phần công việc hay tất cả công việc, mà đã được thực hiện xong tính đến thời điểm báo cáo, mà phần chi phí này mới thực sự làm nên giá trị của phần hay toàn bộ công việc đó. Thường, trong xây dựng công trình, hay các dự án có chi phí được xác định theo định mức thì Giá trị thu được EV được tính bằng tích số giữa khối lượng công việc đã thực sự hoàn thành tính tới thời điểm báo, cáo với định mức hao phí nguồn lực (công lao động và nguyên vật liệu) cho từng công việc.

**Ngân quỹ dự kiến tới thời điểm hoàn thành BAC** (Budget at completion-BAC): Tổng giá trị dự kiến (**PV** hoặc **BCWS**) tính tới thời điểm kết thúc dự án. Nếu một dự án có một Dự phòng phí (Dự trữ quản lý **MR**-Management Reserve), nó thường được thêm vào Ngân quỹ dự kiến tới thời điểm hoàn thành-BAC.

PV = BAC \* % của công việc dự kiến (đạt được theo kế hoạch đến thời điểm theo dõi).

EV = BAC \* % của công việc thực tế (đạt được đến thời điểm theo dõi).

**Chênh lệch chi phí do thay đổi tiến độ SV (SV-Schedule variance)**

**SV** = **EV** – **PV** = BCWP – BCWS, lớn hơn 0 là tốt (trước thời hạn)

**Chỉ số tiến độ thực hiện SPI (SPI-Schedule performance index)**

SPI = EV/PV = BCWP/BCWS, lớn hơn 1 là tốt (trước thời hạn, tức là vượt tiến độ)

**Chêch lệch chi phí CV** (Cost variance) (tức là Tổng chi phí do lệch kế hoạch)

**CV** = **EV** – **AC** = **BCWP** - **ACWP**, lớn hơn 0 là tốt (trong vòng ngân quỹ)

**CV** = (EV - PV) + (PV - AC) = **SV** + **CV**1 = **SV** + **BV**

**Chỉ số chi phí thực hiện CPI** (Cost Performance Index-CPI)

CPI = EV / AC = BCWP / ACWP, lớn hơn 1 là tốt (trong vòng ngân quỹ)

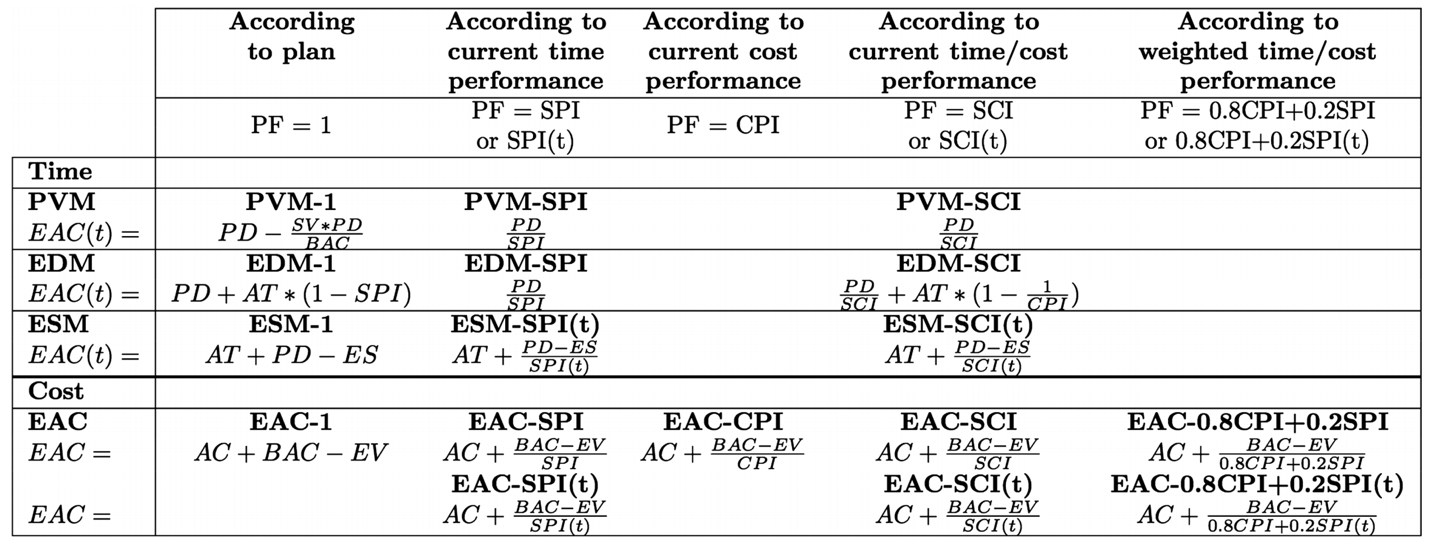
CPI < 1 có nghĩa là chi phí hoàn tất công việc cao hơn so với kế hoạch (vượt ngân sách, có hại)

CPI = 1 có nghĩa là chi phí hoàn thành công việc đúng kế hoạch (có lợi)

CPI > 1 có nghĩa là chi phí hoàn tất các công việc ít hơn dự kiến (dưới ngân sách, tốt nhưng đôi khi có hại).

### Sử dụng kỹ thuật quản trị giá trị thu được trong việc dự đoán thời gian hoàn thành và chi phí dự án

Các phương pháp dự đoán thời gian và chi phí sẽ được tóm tăt trong bảng dưới đây:



Hình 1: Sử dụng EVM dự đoán thời gian và chi phí hoàn thành dự án

**PF** ở trên có ý nghĩa sau đây:

* **PF = 1**: hiệu suất tương đương với kế hoạch đề ra.
* **PF = SPI hoặc SPI(t)**: hiệu suất dựa trên hiệu suất về thời gian của dự án.
* **PF = CPI**: hiệu suất dựa trên hiệu suất về chi phí của dự án.
* **PF = SCI hoặc SCI(t)**: hiệu suất dựa trên hiệu suất về thời gian / chi phí của dự án.
* **PF = 0.8SCI + 0.2SPI hoặc 0.8SCI + 0.2SPI(t):** hiệu suất dựa trên hiệu suất về cả chi phí và thời gian, trong đó chi phí chiếm phần lớn ảnh hưởng (80%).

**Dự đoán về thời gian**

Về dự đoán về thời gian, chúng ta sẽ có 3 phương pháp chính để dự đoán:

- PVM: phương pháp này sử dụng PD (Planned Duration) để tính toán. Kết quả thu được phụ thuộc vào thời gian dự kiến của dự án.

- EDM: phương pháp này sử dụng AT (Actual Time) để tính toán. Do đó, kết quả thu được sẽ phụ thuộc vào hiệu suất tính đến thời điểm hiện tại.

- ESM: sử dụng hiệu suất thực tế tới thời điểm hiện tại để có thể dự đoán thời gian của dự án. Do đó sẽ bị ảnh hưởng rất nhiều về hiệu suất (cả giá trị thu được và cả thời gian đã sử dụng).

**Dự đoán về chi phí**

Về dự đoán chi phí, có 2 cách tính chi phí, và mỗi cách sẽ tùy theo PF chúng ta đã chọn.

- Cách đầu tiên (không tính thời gian hiện tại): chi phí cuối cùng sẽ được tính dựa trên giá trị thu được hiện tại, chi phí hiện tại và tổng chi phí ngân sách.

- Cách thứ hai (có tính thời gian hiện tại): chi phí cuối cùng sẽ được tính bằng cách sử dụng các yếu tố gán liền với mốc thời gian hiện tại.

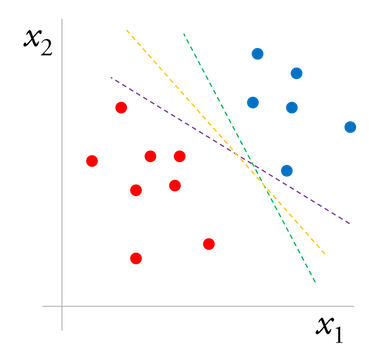
### Ý nghĩa của EVM

EVM cung cấp cho đội ngũ quản lý các dữ liệu đáng tin cậy để phân tích sâu hơn nhằm đưa ra một dấu hiệu cảnh báo sớm về chi phí, thời gian tiến độ của dự án. EVM đo lường các giá trị chi phí theo kế hoạch gốc PV. EVM được coi là cần thiết cho những hành động chủ động và những phân tích quản lý rủi ro.

## Phương pháp SVM

### Support Vector Machine (SVM)

Support Vector Machine (SVM) là một thuật toán thuộc nhóm Supervised Learning dùng để phân chia dữ liệu thành các nhóm riêng biệt.

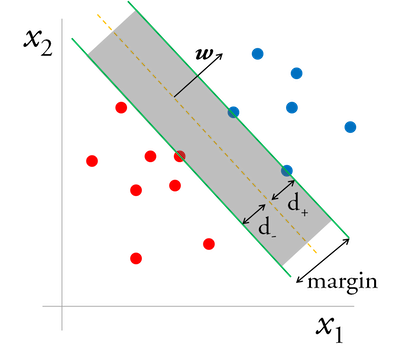
Ý tưởng cơ bản của SVM là tìm ra một siêu phẳng (hyperplane), phân tách không gian dữ liệu thành các miền khác nhau, từ đó phân loại chúng theo mong muốn người sử dụng.

Hình 2: Classify SVM

Giả có bộ data gồm các điểm xanh và đỏ đặt trên cùng một mặt phẳng. Thuật toan SVM có thể tìm được đường thẳng để phân chia riêng biệt các bộ điểm xanh và đỏ.

Tuy nhiên, số lượng các siêu phẳng tìm được là vô hạn do đó bài toán đặt ra là tìm siêu phẳng tối ưu nhất hay tìm siêu phẳng sao cho 2 lớp dữ liệu nằm cách xa nhau và cách xa mặt phẳng nhất.

Margin là khoảng cách giữa siêu phẳng đến 2 điểm dữ liệu gần nhất tương ứng với 2 phân lớp. SVM cố gắng tối ưu thuật toán bằng cách giá trị margin này lớn nhất có thể, từ đó tìm ra siêu phẳng đẹp nhất để phân 2 lớp dữ liệu.

Bài toán của trở thành tìm ra 2 đường biên của 2 lớp dữ liệu sao cho khoảng cách giữa 2 đường này là lớn nhất. Đường biên của lớp xanh sẽ đi qua một (hoặc một vài) điểm xanh. Đường biên của lớp đỏ sẽ đi qua một (hoặc một vài) điểm đỏ. Các điểm xanh, đỏ nằm trên 2 đường biên được gọi là các support vector, vì chúng có nhiệm vụ support để tìm ra siêu phẳng.

Hình 3: Margin SVM

Siêu phẳng cần tìm, hay là siêu phẳng tối ưu (Optimal Hyperplane) sẽ được biểu diễn bằng hàm số <w.x> = b, với:

* w là vector pháp tuyến của hyperplane (không nhất thiết phải chuẩn hóa)
* x là một vector số thực (chiều không gian p)
* <w.x> là tích vô hướng của 2 vector
* b là khoảng cách giữa gốc tọa độ và siêu phẳng, nhân với độ dài vector pháp tuyến w

Giả sử ta có bài toán phân loại các điểm dữ liệu đỏ và xanh như trên hình. Các điểm đỏ có nhãn là 1 (điểm dương), các điểm xanh có nhãn là -1 (điểm âm).

Hyperplane tách 2 lớp dữ liệu H0 thỏa mãn <w.x> + b = 0, tạo ra 2 nửa không gian:

* Không gian các lớp dữ liệu âm xi thỏa mãn <w.xi> + b <= -1
* Không gian các lớp dữ liệu dương xj thỏa mãn <w.xj> + b >= 1

Tiếp theo, ta chọn 2 siêu phẳng lề H1 và H2, đi qua 2 điểm lần lượt âm và dương có khoảng cách gần nhất đến H0, với H1 và H2 cùng song song với H0. Ta có phương trình 2 siêu phẳng H1 và H2:

* H1: <w.xi> + b = -1
* H2: <w.xj> + b = 1

Khoảng cách từ H1 đến H0 là d-, khoảng cách từ H2 đến H0 là d+.

Ta gọi thuộc tính m = d- + d+ là mức lề (margin).

Lý thuyết học máy đã chỉ ra rằng, siêu phẳng với mức lề lớn nhất là siêu phẳng tối ưu, giúp cực tiểu hóa sỗ lỗi có thể mắc phải.

Nếu dữ liệu có thể được tách ra 1 cách tuyến tính (linearly seperable), không gặp một lỗi nào, thì thuật toán được đặt tên là Hard-Margin SVM. Trong một số trường hợp điều này không thể xảy ra (ví dụ như dữ liệu không hoàn toàn tách rời tuyến tính, có nhiễu….), thuật toán Soft-Margin SVM đã được đề xuất để khắc phục điểm yếu của Hard-Margin.

Bài toán Hard-Margin SVM được phát biểu như sau:

Tìm ||W|| tối thiểu để thỏa mãn: A picture containing object, clock

Description automatically generated

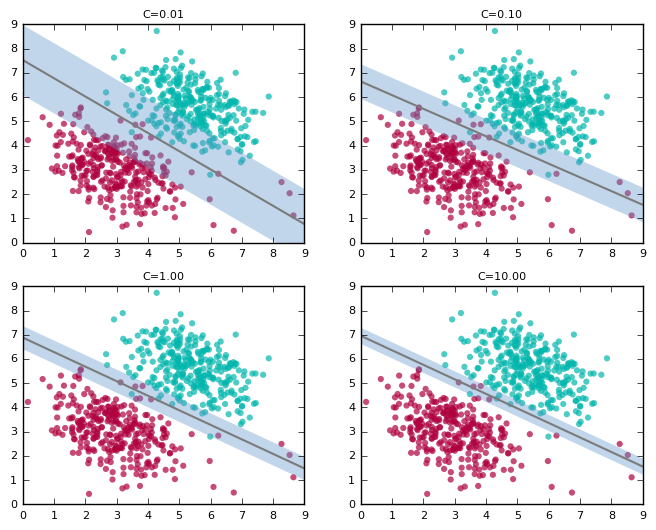
Bài toán Soft-Margin SVM được phát biểu như sau:

Tìm  tối thiểu để thỏa mãn A picture containing object, clock

Description automatically generated

Yêu cầu đặt ra ở đây là tìm ra siêu mặt phẳng tối ưu, trong số lượng siêu mặt phẳng rất lớn.

Tuy nhiên khi tập dữ liệu không phân biệt tuyến tính hoàn toàn, và trong thực tế, dữ liệu cũng không phân biệt tuyến tính hoàn toàn. Nếu sử dụng SVM phân lớp chắc chắn sẽ không thể tách được tập dữ liệu. Dù vậy, không thể bỏ hoàn toàn phương pháp này, bởi vì có lỗi có thể chấp nhận được. Tham số C là tham số để điều chỉnh mức độ bỏ qua các trường hợp phân loại nhầm. Tham số càng lớn margin càng bé, tăng độ chính xác và giảm lỗi khi phân lớp.

Khi tăng C, độ rộng margin giảm, điều này đi ngược với mục đích ban đầu của SVM là tìm cách tối ưu cho độ rộng của margin lớn nhất. Vì vậy cần cân đối giữa 2 tiêu chí này.

Hình 4: Tham số C trong SVM

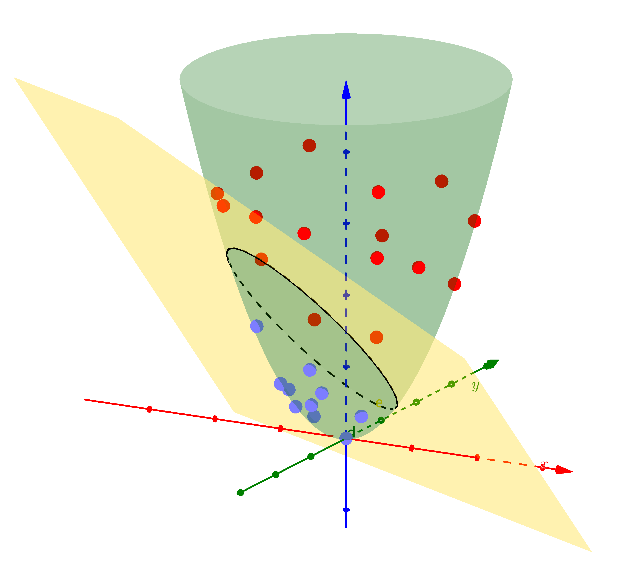
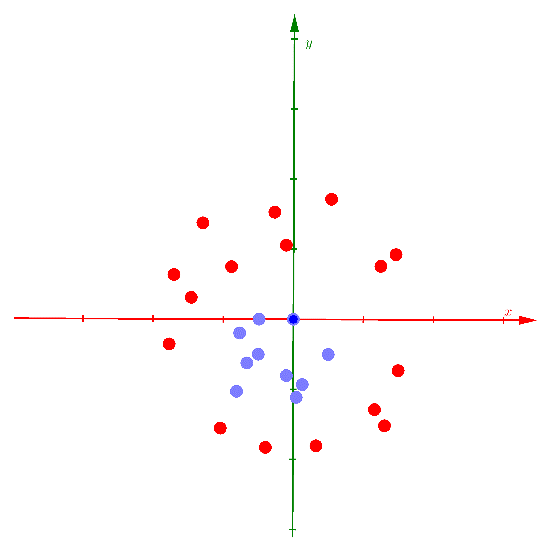
### Kernel trong SVM – RBF Kernel

Khi tập dữ liệu không phân biệt tuyến tính, hay các dữ liệu không phân biệt rõ ràng, nếu chỉ sử dụng SVM thì không thể phân tách được. Do đó, cần đưa tập dữ liệu lên chiều không gian cao hơn, sao cho khi ở không gian mới này, tập dữ liệu trở nên phân biệt tuyến tính hoặc gần phân biệt tuyến tính. SVM thường sử dụng các kernel (hạt nhân) để thực hiện phân tách dữ liệu.

Ví dụ với việc biến dữ liệu không phân biệt tuyến tính trong không gian hai chiều thành phân biệt tuyến tính trong không gian ba chiều bằng cách thêm một chiều mới:

Có 4 loại kernel thông dụng: linear, poly, rbf, sigmoid. Trong đó, rbf được sử dụng nhiều nhất.

Hình 5: Kernel



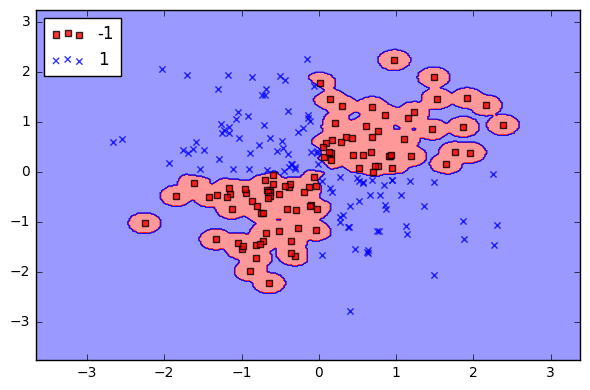
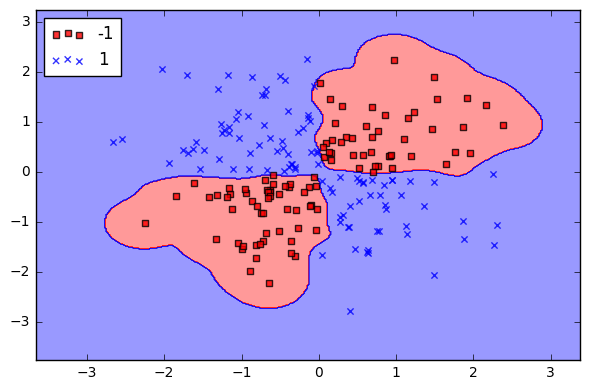
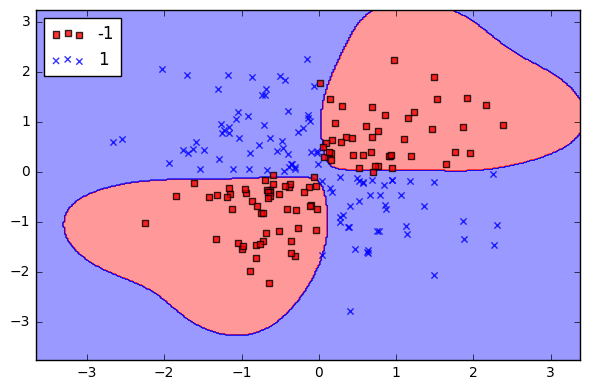
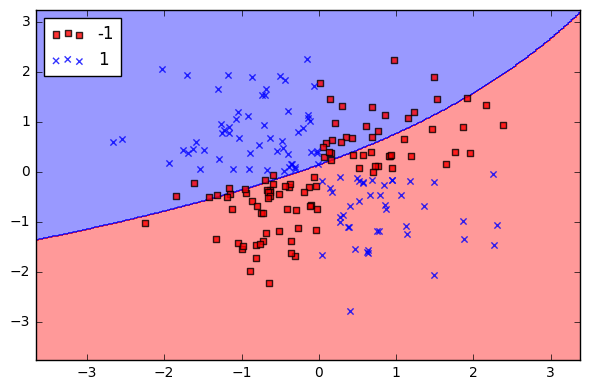
\*RBF kernel:

Radial Basic Function (RBF) kernel hay Gaussian kernel được sử dụng nhiều nhất trong thực tế. Nó được định nghĩa bởi:

RBF là phép chiếu vào vô hạn tập không gian. Khi sử dụng RBF cần thiết lập thông số gamma.

Gamma là tham số điều chỉnh mức độ ảnh hưởng của dữ liệu tới dữ liệu chung. Khi gamma càng lớn, khoảng cách tới siêu phẳng phân tách càng bé.

Với cùng tham số C và thay đổi tham số gamma:



Gamma = 0.1

Gamma = 1

Gamma = 10

Gamma = 100

Hình 6: Tham số gamma trong RBF Kernel

### Support Vector Regression (SVR)

Thuật toán SVR khác các mô hình thuật toán hồi quy khác. SVR sử dụng SVM (thuật toán phân lớp) để dự đoán giá trị biến cần. Trong khi các mô hình hồi quy khác cố gắng làm giá trị dự đoán gần với giá trị thưc, SVR làm giá trị dự đoán phù hợp với mức lỗi chấp nhận được.

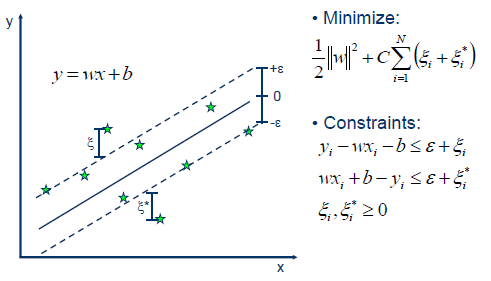
Trong thuật toán này, SVM có chức năng phân loại các kết quả dự đoán thành 2 loại: những kết quả vượt ra ngoài ranh giới lỗi, và những kết quả trong ngưỡng chấp nhận được. SVR có thêm một tham số là epsilon. Epsilon là giới hạn mức độ lỗi chấp nhận được của dự đoán.

Siêu phẳng trong SVR sử dụng để đưa ra các dự đoán

Ranh giới lỗi là 2 siêu phẳng song song về 2 phía của siêu phẳng dự đoán với khoảng cách giá trị chấp nhận lỗi epsilon. 2 siêu phẳng này tạo ra 1 margin.

2 ranh giới lỗi điều chỉnh để chứa được nhiều bộ dữ liệu nhất có thể. Độ rộng giữa 2 ranh giới chính do epsilon quyết định.

Ví dụ một mô hình SVR trên tọa độ 2 chiều:



Hình 7: SVR

Đường nét liền là siêu phẳng sử dụng để đưa ra dự đoán, còn 2 đường nét đứt là siêu phẳng biên. Ở đây chỉ lấy các điểm nằm trong vùng giới hạn của hai biên, vẫn có những điểm ở phía ngoài nhưng ta bỏ qua.

# ÁP DỤNG THUẬT TOÁN SVR ĐỂ ƯỚC LƯỢNG THỜI GIAN VÀ CHI PHÍ HOÀN THÀNH DỰ ÁN PHẦN MỀM

## Bài toán đánh giá mức độ hoàn thành dự án phần mềm

Mục đích chính khi giải quyết bài toán, với 1 lượng dữ liệu cụ thể nhất định tịa một thời điểm khi thực hiện dự án, mô hình sau khi xây dựng đưa ra được con số dự đoán chi phí, thời gian hoàn thành dự án.

Như đã trình bày ở mục 2.2.2, PV (chi phí lên kế hoạch), AC (chi phí bỏ ra tính đến hiện tại), EV (giá trị thu được tính đến hiện tại) là 3 chỉ số quan trọng nhất trong quản lý dự án phần mềm. Từ 3 chỉ số này có thể đánh giá được chất lượng dự án phần mềm, cũng như có thể tính toán trước được chi phí và thời gian hoàn thành dự án.

Đây là bài toán dự đoán, có tính hồi quy nên có thể sử dụng phương pháp học máy để đưa ra dự đoán đánh giá.

Giống như phương pháp truyền thống EVM kết hợp mô hình tăng trường đưa ra dự đoán, trong xây dựng mô hình SVR, đề tài cũng sử dụng chỉ số AC, EV cùng với AT (thời gian hiện tại, các mốc thời gian) để xây dựng mô hình dự đoán chi phí, thời gian hoàn thành dự án phần mềm.

Mô hình xây dựng bằng cách train qua nhiều bộ dữ liệu quản lý phần mềm, nhiều dự án phần mềm có các chỉ số tương ứng, và từ đó đưa ra đánh giá cho bộ dữ liệu cần.

### Dự đoán thời gian hoàn thành dự án

Kế thừa từ chương trình của K61, khi xây dựng mô hình dự đoán thời gian, trong đề tài này cũng sử dụng 3 thông số:

* Input:
  + AT: thời gian hiện tại
  + XT: tiến độ thực hiện dự án, tính bằng tỉ lệ EV tại thời điểm AT và TOTAL\_EV là tổng giá trị thu được ước tính
* Output:
  + TOTAL\_AT: tổng thời gian thực hiện dự án

Và từ đó sử dụng model để dự đoán TOTAL\_AT trong các dự án khác

### Dự đoán chi phí hoàn thành dự án

Giống với dự đoán thời gian, khi xây dựng mô hình cũng sử dụng thông số XT, tuy nhiên thay vì sử dụng dữ liệu liên quan tới AT, chương trình sử dụng AC.

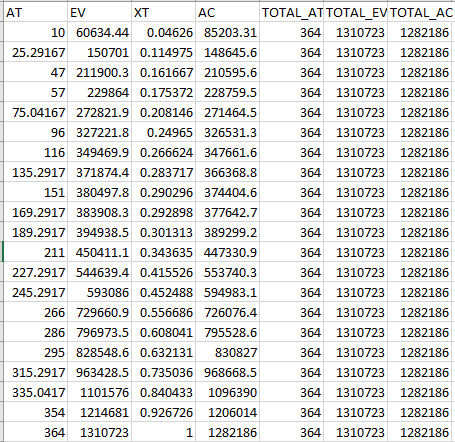
* Input:
  + XT
  + AC: chi phí bỏ ra tính tới thời điểm hiện tại
* Output:
  + TOTAL\_AC: tổng chi phí của dự án

## Xây dựng tập dữ liệu

Để xây dựng được một mô hình học máy, train model điều quan trọng nhất là dữ liệu, dữ liệu càng nhiều càng đa dạng, mô hình càng chính xác.

### Tập dữ liệu kế thừa từ các khóa trước

Các bộ dữ liệu này đặt trong thư mục dataset/ của project. Dữ liệu đã được nhóm nghiên cứu trước thu thập, xử lí và có cấu trúc như sau:



Hình 8: Dữ liệu sau xử lí

Trong đó:

* AT: Mốc thời gian thực hiện dự án (đo bằng ngày)
* PV: chi phí được lên kế hoạch (đo bằng tiền).
* XT: tiến độ thực hiện dự án (XT= EVcurrent/TOTAL\_EV).
* AC: chi phi bỏ ra tính đến hiện tại (đo bằng tiền).
* TOTAL\_AT: tổng số thời gian thực hiện dự án.
* TOTAL\_PV: tổng chi phí hoàn thành dự án được lên kế hoạch.
* TOTAL\_EV: tổng chi phí ước lượng đạt được của dự án.
* TOTAL\_AV: tổng chi phí bỏ ra để hoàn thành dự án.

### Thu thập thêm dữ liệu

Dữ liệu của đề tài bọn em lấy từ trang web:

<http://www.projectmanagement.ugent.be/research/data/realdata>

Trang web có 133 bộ dữ liệu về dự án. Tuy nhiên cũng có một số bộ dữ liệu bị trùng với khóa trước, những bộ dữ liệu trùng được loại bỏ.

## Xây dựng mô hình SVR

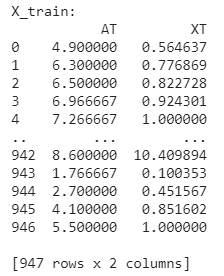
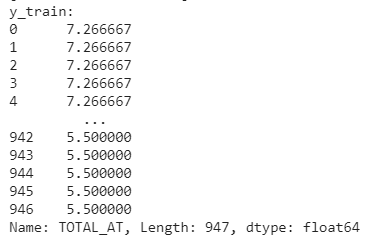
### Dữ liệu huấn luyện, dữ liệu test

Vì đề tài thực hiện trên môi trường python, nên để xây dựng mô hình SVR nhóm em dùng thư viện có sẵn của python hỗ trợ các thuật toán học máy là sklearn.

Trên tổng số 133 file dữ liệu sử dụng huấn luyện cho mô hình, chia ngẫu nhiên để train test theo tỷ lệ 75-25. 75% file sử dụng để huấn luyện xây dựng model.

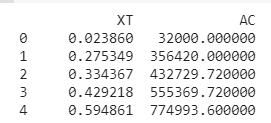
Dữ liệu tổ chức thành 2 file riêng biệt, train và test. Chương trình đọc dữ liệu cần vào qua thư viện pandas. Ví dụ dữ liệu train test dự đoán thời gian thực hiện có dạng như sau:

Hình 9: Dữ liệu input của chương trình



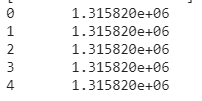
Với model dữ đoán thời gian, dữ liệu gồm 3 cột AT, XT và TOTAL\_AT với x\_train là AT, XT y\_train là TOTAL\_AT. Model dữ đoán chi phí, dữ liệu gồm 3 cột AC, XT và TOTAL\_AC.

Tổ chức dữ liệu X\_train với dataFrame.iloc của pandas



Hình 10: x\_train

Và y\_train chỉ 1 cột duy nhất:



Hình 11: y\_train

### Lựa chọn tham số C, Gamma

Như đã trình bày ở trên, khi xây dựng mô hình SVR có 2 chỉ số cần quan tâm là C, Gamma (không tính tới kernel vì đã chọn RBF). Do đó cần có cách lựa chọn 2 chỉ số này để chương trình tối ưu nhất. Giải pháp là sử dụng GridSearchCV ước lượng tham số xác thực chéo.

Thay vì tự chọn chính xác 1 tham số cho mô hình và thực hiện nhiều lần, GridSearchCV sử dụng một list các tham số và tự thực hiện train với các tham số. Tham số giúp model đưa ra kết quả dự đoán chính xác nhất được sử dụng.

### Đánh giá kết quả dự đoán

Đề tài sử dụng 4 cách đưa ra đánh giá kết quả dự đoán để đánh giá mô hình, sử dụng explained\_variance\_score.

* Tại lần báo khoảng 25%
* Đưa ra các dự đoán từ khi bắt đầu dự án tới mốc 25%
* Tại lần báo cáo 50%
* Đưa ra các dự đoán từ khi bắt đầu dự án tới mốc 50%

# CÀI ĐẶT VÀ ĐÁNH GIÁ THỰC NGHIỆM

## Môi trường thực nghiệm

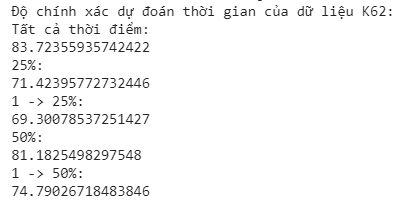
Thực hiện đề tài, nhóm em triển khai trên google colab, với đề tài, chương trình sử dụng các thư viện của python:

* Glob: đọc file trong thư mục
* Pandas, numpy: 2 thư viện xử lí dữ liệu
* Sklearn: chứa thư viện học máy SVR, gridsearch và hàm đánh giá explained\_variance\_score
* Joblib: lưu trữ mô hình
* Các thư viện toán học: random, floor…

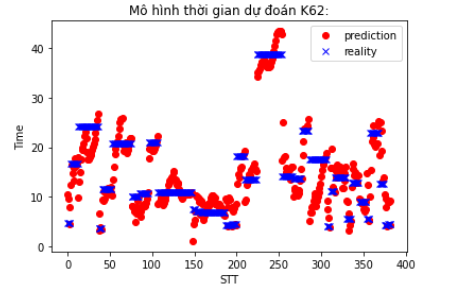
## Kết quả thực nghiệm

Kết quả chi tiết ở phần phụ lục

* Mô hình dự đoán thời gian hoàn thành dự án:

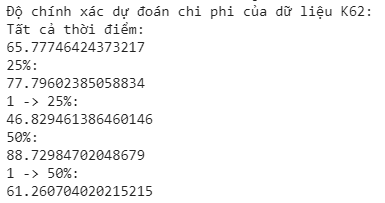


Hình 12: Kết quả dự đoán thời gian hoàn thành dự án tại các mốc thời gian K62

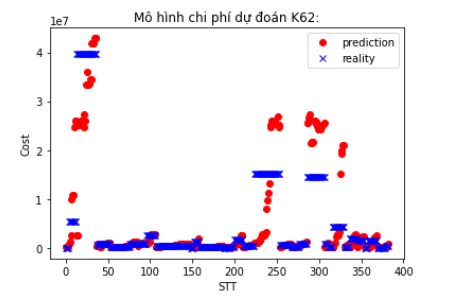


Hình 13: Biểu đồ kết quả dự đoán và kết quả thực tế thời gian hoàn thành dự án

* Mô hình dự đoán chi phí hoàn thành dự án



Hình 14: Kết quả dự đoán chi phí hoàn thành dự án tại các mốc thời gian K62

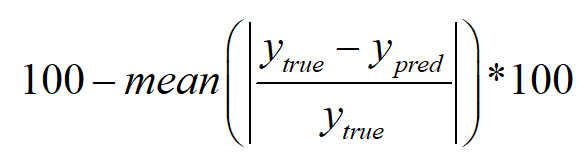


Hình 15: Biểu đồ kết quả dự đoán và kết quả thực tế chi phí hoàn thành dự án

## Đánh giá mô hình

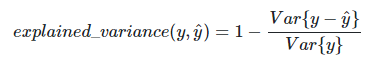
Với mô hình dự đoán thời gian hoàn thành dự án, ở các giai đoạn, kết quả dự đoán so với thực tế đúng xấp xỉ 70%. Đây là kết quả chấp nhận được và có thể sử dụng trong thực tế.

Metrics được sử dụng là mean\_absolute\_percentage\_error được tính theo công thức:



Đây là metric được sử dụng từ K61 bọn em sử dụng để so sánh kết quả, tuy nhiên trong trường hợp ypred > 2\*ytrue thì giá trị của mean\_absolute\_percentage\_error sẽ < 0. Trong một số trường hợp bọn em test và được kết quả mean\_absolute\_percentage\_error < 0

Để khắc phục nhược điểm này, bọn em quyết định sử dụng thêm Explained variance score để đánh giá. Đây là metric thường được sử dụng trong các bài toán hồi quy:



Mô hình được đánh giá theo 4 tiêu chí 25%, 0->25%, 50% và 0->50% dự án.

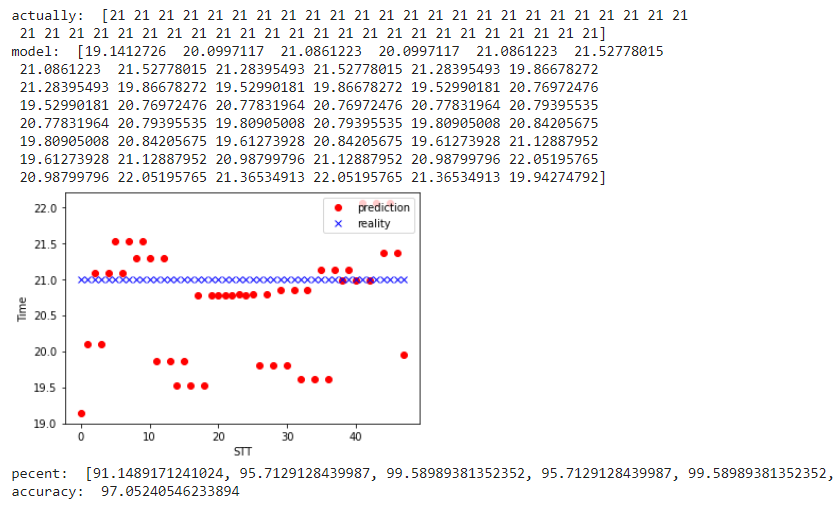
Với 25 và 50, mỗi dữ liệu dự án sẽ lấy một bán ghi bổ sung vào tập train và test. Trong thực tế, đây là thời điểm đự án đạt đến 25% hoặc 50% theo kế hoạch. Nhóm phát triển dự án đánh giá lại những gì mình đã làm được, dự đoán thời điểm, chi phí của dự án.

Tại mốc 25% dựa vào dự báo đưa ra bản kế hoạch dài hạn, đánh giá rủi ro sớm và đề ra kế hoạch tiếp theo (thay đổi nhân sự, tăng giảm nguồn tiền, thay đổi yêu cầu, công nghệ). Thuật toán học máy tỏ ra có ưu thế hơn khi có sự thay đổi mạnh, liên tục như vậy, thay vì dự báo theo lịch sử dự án hiện tại. Sử dụng bộ dữ liệu của các dự án khác sẽ đưa ra cái nhìn tổng quan hơn. Mô hình này có thể được sử dụng như một giải pháp dự đoán thay thế khi đội phát triển dự án muốn thay đổi một cách hoàn toàn 75% kế hoạch còn lại.

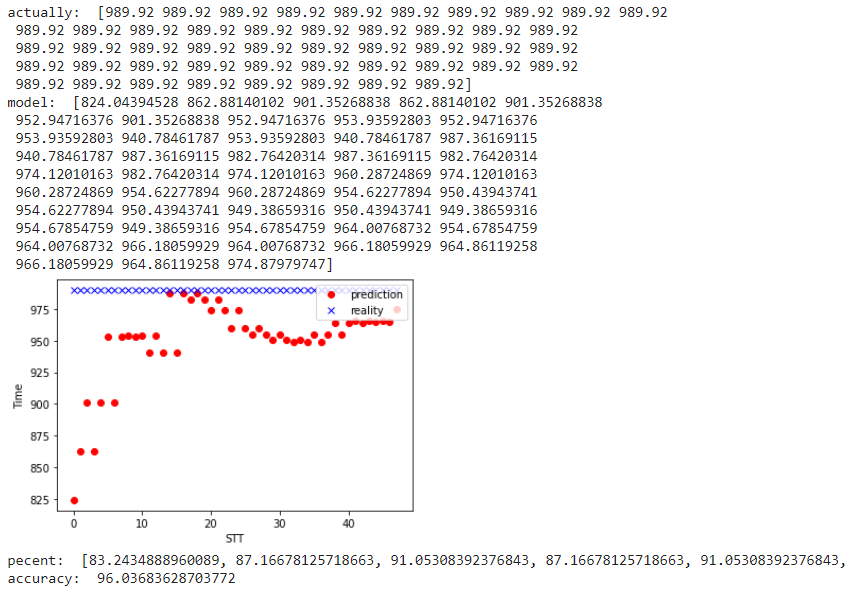
Tại mốc 50% ưu thế này dường như đã bị giảm bớt, khi dự án đã hoàn thành được một mức độ nhất định. Dù rằng tỉ lệ dự đoán cao hơn nhưng không mang lại nhiều ý nghĩa thực tế so với mốc 25%. Trong trường hợp đội phát triển dự án vẫn muốn thay đổi mạnh một vài nhân tố quyết định thời điểm và chi phí hoàn thành dự án điển hình như tuyển thêm senior, cho các thành viên làm thêm giờ hay làm việc tại nhà (một việc khá phổ biến trong đợt dịch NCOV-20). Mô hình dự báo sử dụng thuật toán học máy cũng là một giải pháp đáng lưu ý.

Mốc tất cả sử dụng tất cả các bản ghi của bộ dự án thêm vào tập train, test được sử dụng để so sánh với mô hình của K61.

## So sánh chương trình của K61



Hình 16: Kết quả dự đoán thời gian hoàn thành dự án của K61



Hình 17: Kết quả dự đoán chi phí hoàn thành dự án của K61

Mô hình k61 có thể dự đoán với độ chính xác trung bình hơn 95%. Tuy nhiên, tỷ lệ này chỉ đánh giá trên 1 file test duy nhất. Mô hình của k62 độ chính xác chỉ khoảng 70%, không chính xác bằng mô hình của k61 nếu so sánh tỷ lệ. Tuy nhiên tỷ lệ 70% này là độ chính xác trung bình trên nhiều file kết quả trên 30 bộ test khi áp dụng mô hình để dự đoán. Do đó, mặc dù kết quả dự đoán của k62 thấp hơn nhưng có ý nghĩa và thực tiễn cao hơn.

## So sánh với phương pháp EVM

|  |  |  |
| --- | --- | --- |
|  | Phương pháp SVR | Phương pháp EVM |
| Dự đoán thời điểm hoàn thành (0-25%) | 69.3% | 85.924% |
| Dự đoán thời điểm hoàn thành (0-50%) | 81.18% | 91.214% |
| Dự đoán chi phí hoàn thành (0-25%) | 46.8% | 91.647% |
| Dự đoán chi phí hoàn thành (0-50%) | 88.7% | 93.367% |

# KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN

## Kết luận

Mô hình dự đoán chi phí và thời gian hoàn thành dự án được xây dựng từ hơn 100 file dữ liệu của các dự án phần mềm. Mô hình là sự kết hợp phương pháp EVM và thuật toán học máy SVR.

Giống với mô hình của k61, mô hình k62 có thể đưa ra dự đoán ở bất cứ giai đoạn nào trong dự án. Với đầu vào gồm 3 tham số [XT, AT, AC] mô hình sẽ đưa ra dự đoán ngay tại giai đoạn đó. Đặc biệt tại ở các khoảng thời gian đầu dự án (trước 50% dự án) mô hình đưa ra dự đoán với bộ dữ liệu khá chính xác.

Mô hình K62 được đánh giá trên 40 bộ test và có tỷ lệ chính xác khoảng 70%, đây là tỷ lệ chấp nhận được khi dữ liệu xây dựng mô hình còn ít.

Link:

https://drive.google.com/drive/folders/1PecWK5-yVeUaA1DRJEFTNs0cqQfJgqUM?usp=sharing

## Định hướng phát triển

Điều quan trọng nhất trong đề tài khiến kết quả dự đoán thấp đó là dữ liệu. Mô hình xây dựng chỉ dựa trên 130 bộ dữ liệu như vậy là chưa đủ. Do đó nếu muốn phát triền được mô hình SVR thì cần thêm nhiều dữ liệu nữa để training.

Đây là vấn đề khó khăn nhất mà nhóm gặp phải khi thực hiện đề tài. Nguồn dữ liệu về dự án phần mềm rất ít. Nhóm có tham khảo bộ dữ liệu trên trang quản lý dự án phần mềm <https://www.isbsg.org/> nhưng dữ liệu không theo mô hình cũ AT, EV, AC. Đây là nguồn dữ liệu khổng lồ mà các nhóm nghiên cứu sau có thể cân nhắc sử dụng.

# DANH MỤC TÀI LIỆU THAM KHẢO

1. Wikipedia Article, about Support Vector Machine

< <https://en.wikipedia.org/wiki/Support_vector_machine>>

2. Đỗ Minh Hải, 22/03/2018, Bài viết về SVM

< <https://dominhhai.github.io/vi/2018/03/ml-svm/>>

3. 11/05/2018, Bài viết về SVM

< <https://neralnetwork.wordpress.com/2018/05/11/thuat-toan-support-vector-machine-svm/>>

4. Website cung cấp thông số các dự án thực tế

< <http://www.projectmanagement.ugent.be/research/data/realdata> >

5. Sklearn document

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>

# PHỤ LỤC

Dự đoán thời gian:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dự án | Loại | Thực tế | Dự đoán | Độ chính xác |
| C2016-22.csv | tat ca | 4.679167 | 8.131657 | 26.21569 |
| C2016-22.csv | 25 | 4.679167 | 9.722697 | 0 |
| C2016-22.csv | 1 -> 25 | 4.679167 | 10.56219 | 0 |
| C2016-22.csv | 50 | 4.679167 | 7.831792 | 32.62422 |
| C2016-22.csv | 1 -> 50 | 4.679167 | 10.14244 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-07.csv | tat ca | 16.7125 | 15.77826 | 94.40995 |
| C2015-07.csv | 25 | 16.7125 | 14.86905 | 88.96966 |
| C2015-07.csv | 1 -> 25 | 16.7125 | 14.50772 | 86.80758 |
| C2015-07.csv | 50 | 16.7125 | 15.8508 | 94.84397 |
| C2015-07.csv | 1 -> 50 | 16.7125 | 14.29668 | 85.54483 |
| ----- | ------ | ------ | ------ | ------ |
| C2014-01.csv | tat ca | 24.14583 | 19.73484 | 81.73187 |
| C2014-01.csv | 25 | 24.14583 | 20.74706 | 85.92396 |
| C2014-01.csv | 1 -> 25 | 24.14583 | 15.80655 | 65.46285 |
| C2014-01.csv | 50 | 24.14583 | 17.49562 | 72.45814 |
| C2014-01.csv | 1 -> 50 | 24.14583 | 18.30276 | 75.8009 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-28.csv | tat ca | 3.6125 | 4.262517 | 82.00645 |
| C2016-28.csv | 25 | 3.6125 | 3.295032 | 91.21197 |
| C2016-28.csv | 1 -> 25 | 3.6125 | 6.601349 | 17.26369 |
| C2016-28.csv | 50 | 3.6125 | 3.826594 | 94.07352 |
| C2016-28.csv | 1 -> 50 | 3.6125 | 4.948191 | 63.02586 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-04.csv | tat ca | 11.53333 | 10.67546 | 92.56184 |
| C2016-04.csv | 25 | 11.53333 | 9.677219 | 83.90653 |
| C2016-04.csv | 1 -> 25 | 11.53333 | 9.942321 | 86.2051 |
| C2016-04.csv | 50 | 11.53333 | 11.10163 | 96.25688 |
| C2016-04.csv | 1 -> 50 | 11.53333 | 10.23062 | 88.70476 |
| ----- | ------ | ------ | ------ | ------ |
| C2011-07.csv | tat ca | 20.75 | 20.56054 | 99.08694 |
| C2011-07.csv | 25 | 20.75 | 20.80492 | 99.73533 |
| C2011-07.csv | 1 -> 25 | 20.75 | 15.64641 | 75.4044 |
| C2011-07.csv | 50 | 20.75 | 25.60795 | 76.58817 |
| C2011-07.csv | 1 -> 50 | 20.75 | 19.73829 | 95.12428 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-13.csv | tat ca | 10.07917 | 7.758743 | 76.97802 |
| C2013-13.csv | 25 | 10.07917 | 7.924708 | 78.62464 |
| C2013-13.csv | 1 -> 25 | 10.07917 | 8.152908 | 80.88871 |
| C2013-13.csv | 50 | 10.07917 | 4.986972 | 49.47802 |
| C2013-13.csv | 1 -> 50 | 10.07917 | 7.799239 | 77.3798 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-03.csv | tat ca | 10.73333 | 8.239157 | 76.76233 |
| C2016-03.csv | 25 | 10.73333 | 7.088136 | 66.03853 |
| C2016-03.csv | 1 -> 25 | 10.73333 | 6.712292 | 62.53688 |
| C2016-03.csv | 50 | 10.73333 | 8.485934 | 79.06149 |
| C2016-03.csv | 1 -> 50 | 10.73333 | 7.290235 | 67.92145 |
| ----- | ------ | ------ | ------ | ------ |
| data\_020.csv | tat ca | 21.0667 | 19.93658 | 94.63553 |
| data\_020.csv | 25 | 21.0667 | 20.40135 | 96.84172 |
| data\_020.csv | 1 -> 25 | 21.0667 | 18.70186 | 88.77453 |
| data\_020.csv | 50 | 21.0667 | 19.30005 | 91.61401 |
| data\_020.csv | 1 -> 50 | 21.0667 | 19.39945 | 92.08586 |
| ----- | ------ | ------ | ------ | ------ |
| data\_012.csv | tat ca | 10.8667 | 11.16674 | 97.23891 |
| data\_012.csv | 25 | 10.8667 | 10.63125 | 97.83326 |
| data\_012.csv | 1 -> 25 | 10.8667 | 9.251903 | 85.13995 |
| data\_012.csv | 50 | 10.8667 | 13.49499 | 75.8134 |
| data\_012.csv | 1 -> 50 | 10.8667 | 11.2348 | 96.61257 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-20.csv | tat ca | 7.486111 | 4.441143 | 59.32511 |
| C2015-20.csv | 25 | 7.486111 | 4.428996 | 59.16284 |
| C2015-20.csv | 1 -> 25 | 7.486111 | 1.075265 | 14.36347 |
| C2015-20.csv | 50 | 7.486111 | 5.301096 | 70.81241 |
| C2015-20.csv | 1 -> 50 | 7.486111 | 2.752131 | 36.76315 |
| ----- | ------ | ------ | ------ | ------ |
| data\_017.csv | tat ca | 7.0333 | 9.22485 | 68.84037 |
| data\_017.csv | 25 | 7.0333 | 10.05115 | 57.09204 |
| data\_017.csv | 1 -> 25 | 7.0333 | 10.28197 | 53.8101 |
| data\_017.csv | 50 | 7.0333 | 9.574789 | 63.86491 |
| data\_017.csv | 1 -> 50 | 7.0333 | 10.09024 | 56.53614 |
| ----- | ------ | ------ | ------ | ------ |
| C2014-02.csv | tat ca | 6.845833 | 8.083381 | 81.92261 |
| C2014-02.csv | 25 | 6.845833 | 8.01352 | 82.9431 |
| C2014-02.csv | 1 -> 25 | 6.845833 | 7.051129 | 97.00116 |
| C2014-02.csv | 50 | 6.845833 | 10.20022 | 51.00105 |
| C2014-02.csv | 1 -> 50 | 6.845833 | 8.009207 | 83.0061 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-20.csv | tat ca | 4.2125 | 7.69735 | 17.27359 |
| C2016-20.csv | 25 | 4.2125 | 9.28898 | 0 |
| C2016-20.csv | 1 -> 25 | 4.2125 | 9.689746 | 0 |
| C2016-20.csv | 50 | 4.2125 | 7.858143 | 13.45655 |
| C2016-20.csv | 1 -> 50 | 4.2125 | 9.489363 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-18.csv | tat ca | 4.211111 | 7.75579 | 15.82556 |
| C2016-18.csv | 25 | 4.211111 | 9.508706 | 0 |
| C2016-18.csv | 1 -> 25 | 4.211111 | 9.799461 | 0 |
| C2016-18.csv | 50 | 4.211111 | 7.763862 | 15.63389 |
| C2016-18.csv | 1 -> 50 | 4.211111 | 9.654084 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| Sheet21.csv | tat ca | 4.577778 | 6.317994 | 61.98556 |
| Sheet21.csv | 25 | 4.577778 | 7.13774 | 44.07848 |
| Sheet21.csv | 1 -> 25 | 4.577778 | 7.945606 | 26.43093 |
| Sheet21.csv | 50 | 4.577778 | 7.641282 | 33.0788 |
| Sheet21.csv | 1 -> 50 | 4.577778 | 7.541673 | 35.25471 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-09.csv | tat ca | 18.17917 | 15.32492 | 84.29934 |
| C2013-09.csv | 25 | 18.17917 | 14.48674 | 79.6887 |
| C2013-09.csv | 1 -> 25 | 18.17917 | 10.6229 | 58.43448 |
| C2013-09.csv | 50 | 18.17917 | 15.79276 | 86.87287 |
| C2013-09.csv | 1 -> 50 | 18.17917 | 13.04201 | 71.74152 |
| ----- | ------ | ------ | ------ | ------ |
| Sheet19.csv | tat ca | 13.44444 | 14.09829 | 95.1367 |
| Sheet19.csv | 25 | 13.44444 | 13.03258 | 96.93653 |
| Sheet19.csv | 1 -> 25 | 13.44444 | 10.29389 | 76.56613 |
| Sheet19.csv | 50 | 13.44444 | 15.50975 | 84.63822 |
| Sheet19.csv | 1 -> 50 | 13.44444 | 12.35815 | 91.92016 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-10.csv | tat ca | 38.7125 | 38.53977 | 99.55382 |
| C2013-10.csv | 25 | 38.7125 | 37.05173 | 95.71 |
| C2013-10.csv | 1 -> 25 | 38.7125 | 35.94351 | 92.84731 |
| C2013-10.csv | 50 | 38.7125 | 36.28357 | 93.72573 |
| C2013-10.csv | 1 -> 50 | 38.7125 | 36.37426 | 93.95999 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-08.csv | tat ca | 14.13472 | 16.49253 | 83.31907 |
| C2013-08.csv | 25 | 14.13472 | 18.21227 | 71.15227 |
| C2013-08.csv | 1 -> 25 | 14.13472 | 20.1082 | 57.73901 |
| C2013-08.csv | 50 | 14.13472 | 15.04305 | 93.57381 |
| C2013-08.csv | 1 -> 50 | 14.13472 | 18.23263 | 71.00818 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-06.csv | tat ca | 13.67917 | 11.63432 | 85.05139 |
| C2015-06.csv | 25 | 13.67917 | 11.73738 | 85.80478 |
| C2015-06.csv | 1 -> 25 | 13.67917 | 10.65111 | 77.86376 |
| C2015-06.csv | 50 | 13.67917 | 10.2062 | 74.61125 |
| C2015-06.csv | 1 -> 50 | 13.67917 | 11.01418 | 80.5179 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-02.csv | tat ca | 23.34583 | 22.09495 | 94.64192 |
| C2015-02.csv | 25 | 23.34583 | 21.1201 | 90.46623 |
| C2015-02.csv | 1 -> 25 | 23.34583 | 16.08449 | 68.89661 |
| C2015-02.csv | 50 | 23.34583 | 23.71302 | 98.42718 |
| C2015-02.csv | 1 -> 50 | 23.34583 | 19.26538 | 82.52172 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-10.csv | tat ca | 17.58056 | 11.68013 | 66.43778 |
| C2016-10.csv | 25 | 17.58056 | 9.154377 | 52.07104 |
| C2016-10.csv | 1 -> 25 | 17.58056 | 6.99823 | 39.80665 |
| C2016-10.csv | 50 | 17.58056 | 10.15602 | 57.76847 |
| C2016-10.csv | 1 -> 50 | 17.58056 | 8.224443 | 46.78147 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-27.csv | tat ca | 4.045833 | 6.58556 | 37.22612 |
| C2016-27.csv | 25 | 4.045833 | 7.520165 | 14.12568 |
| C2016-27.csv | 1 -> 25 | 4.045833 | 9.27975 | 0 |
| C2016-27.csv | 50 | 4.045833 | 5.759255 | 57.64972 |
| C2016-27.csv | 1 -> 50 | 4.045833 | 8.399958 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-17.csv | tat ca | 11.08611 | 14.94769 | 65.1674 |
| C2015-17.csv | 25 | 11.08611 | 19.65589 | 22.69807 |
| C2015-17.csv | 1 -> 25 | 11.08611 | 15.92069 | 56.39067 |
| C2015-17.csv | 50 | 11.08611 | 15.91925 | 56.40362 |
| C2015-17.csv | 1 -> 50 | 11.08611 | 17.78829 | 39.54437 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-06.csv | tat ca | 14 | 14.36654 | 97.38183 |
| C2016-06.csv | 25 | 14 | 16.4916 | 82.20283 |
| C2016-06.csv | 1 -> 25 | 14 | 13.28167 | 94.86905 |
| C2016-06.csv | 50 | 14 | 15.70503 | 87.8212 |
| C2016-06.csv | 1 -> 50 | 14 | 14.66124 | 95.27686 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-15.csv | tat ca | 5.513889 | 6.244156 | 86.75586 |
| C2013-15.csv | 25 | 5.513889 | 9.04574 | 35.94628 |
| C2013-15.csv | 1 -> 25 | 5.513889 | 9.233506 | 32.54095 |
| C2013-15.csv | 50 | 5.513889 | 3.182896 | 57.72506 |
| C2013-15.csv | 1 -> 50 | 5.513889 | 8.293901 | 49.58165 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-03.csv | tat ca | 12.9125 | 14.19618 | 90.05862 |
| C2015-03.csv | 25 | 12.9125 | 15.64925 | 78.80545 |
| C2015-03.csv | 1 -> 25 | 12.9125 | 14.31034 | 89.17448 |
| C2015-03.csv | 50 | 12.9125 | 14.89459 | 84.64987 |
| C2015-03.csv | 1 -> 50 | 12.9125 | 14.65248 | 86.5248 |
| ----- | ------ | ------ | ------ | ------ |
| data\_18.csv | tat ca | 9.033333 | 8.449897 | 93.54129 |
| data\_18.csv | 25 | 9.033333 | 8.739234 | 96.74429 |
| data\_18.csv | 1 -> 25 | 9.033333 | 9.509252 | 94.73153 |
| data\_18.csv | 50 | 9.033333 | 8.594918 | 95.1467 |
| data\_18.csv | 1 -> 50 | 9.033333 | 9.11315 | 99.11642 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-23.csv | tat ca | 5.6 | 9.447376 | 31.29686 |
| C2015-23.csv | 25 | 5.6 | 12.48301 | 0 |
| C2015-23.csv | 1 -> 25 | 5.6 | 14.80316 | 0 |
| C2015-23.csv | 50 | 5.6 | 5.283558 | 94.34924 |
| C2015-23.csv | 1 -> 50 | 5.6 | 13.64309 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2019-6.csv | tat ca | 23 | 20.44594 | 88.89541 |
| C2019-6.csv | 25 | 23 | 21.25013 | 92.39189 |
| C2019-6.csv | 1 -> 25 | 23 | 12.71785 | 55.29498 |
| C2019-6.csv | 50 | 23 | 20.53311 | 89.27438 |
| C2019-6.csv | 1 -> 50 | 23 | 18.61853 | 80.95014 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-16.csv | tat ca | 12.6 | 18.84946 | 50.40113 |
| C2015-16.csv | 25 | 12.6 | 25.08725 | 0.894865 |
| C2015-16.csv | 1 -> 25 | 12.6 | 18.34502 | 54.40461 |
| C2015-16.csv | 50 | 12.6 | 19.272 | 47.0476 |
| C2015-16.csv | 1 -> 50 | 12.6 | 22.2721 | 23.23732 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-25.csv | tat ca | 4.2125 | 8.017347 | 9.677233 |
| C2016-25.csv | 25 | 4.2125 | 9.572915 | 0 |
| C2016-25.csv | 1 -> 25 | 4.2125 | 10.59036 | 0 |
| C2016-25.csv | 50 | 4.2125 | 7.953584 | 11.19089 |
| C2016-25.csv | 1 -> 50 | 4.2125 | 10.08164 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-32.csv | tat ca | 4.479167 | 7.890414 | 23.84192 |
| C2016-32.csv | 25 | 4.479167 | 9.078739 | 0 |
| C2016-32.csv | 1 -> 25 | 4.479167 | 8.963747 | 0 |
| C2016-32.csv | 50 | 4.479167 | 9.30197 | 0 |
| C2016-32.csv | 1 -> 50 | 4.479167 | 9.021243 | 0 |
| ----- | ------ | ------ | ------ | ------ |

Dự đoán chi phí:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mã | Loại | Thực tế | Dự đoán | Chính xác |
| C2016-22.csv | tat ca | 57822.4 | 162869.7 | 0 |
| C2016-22.csv | 25 | 57822.4 | 151242.6 | 0 |
| C2016-22.csv | 1 -> 25 | 57822.4 | 141232.6 | 0 |
| C2016-22.csv | 50 | 57822.4 | 168111.8 | 0 |
| C2016-22.csv | 1 -> 50 | 57822.4 | 146237.6 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-07.csv | tat ca | 5414544 | 12608311 | 0 |
| C2015-07.csv | 25 | 5414544 | 2681402 | 49.52221 |
| C2015-07.csv | 1 -> 25 | 5414544 | 1136493 | 20.98964 |
| C2015-07.csv | 50 | 5414544 | 10890787 | 0 |
| C2015-07.csv | 1 -> 50 | 5414544 | 3736473 | 69.00809 |
| ----- | ------ | ------ | ------ | ------ |
| C2014-01.csv | tat ca | 39777643 | 30427553 | 76.49411 |
| C2014-01.csv | 25 | 39777643 | 26087596 | 65.58356 |
| C2014-01.csv | 1 -> 25 | 39777643 | 17818401 | 44.79501 |
| C2014-01.csv | 50 | 39777643 | 36009869 | 90.52791 |
| C2014-01.csv | 1 -> 50 | 39777643 | 22547344 | 56.68346 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-28.csv | tat ca | 586086.8 | 761795.5 | 70.02003 |
| C2016-28.csv | 25 | 586086.8 | 860989.9 | 53.09517 |
| C2016-28.csv | 1 -> 25 | 586086.8 | 379590.6 | 64.76696 |
| C2016-28.csv | 50 | 586086.8 | 873384.1 | 50.98043 |
| C2016-28.csv | 1 -> 50 | 586086.8 | 620290.3 | 94.1641 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-04.csv | tat ca | 906253.9 | 818736.3 | 90.34294 |
| C2016-04.csv | 25 | 906253.9 | 825998.5 | 91.14428 |
| C2016-04.csv | 1 -> 25 | 906253.9 | 339843.2 | 37.49978 |
| C2016-04.csv | 50 | 906253.9 | 927683.2 | 97.63539 |
| C2016-04.csv | 1 -> 50 | 906253.9 | 598916.1 | 66.08702 |
| ----- | ------ | ------ | ------ | ------ |
| C2011-07.csv | tat ca | 191065.1 | 240038 | 74.36846 |
| C2011-07.csv | 25 | 191065.1 | 148743.2 | 77.84953 |
| C2011-07.csv | 1 -> 25 | 191065.1 | 142037.6 | 74.33991 |
| C2011-07.csv | 50 | 191065.1 | 158397.1 | 82.90219 |
| C2011-07.csv | 1 -> 50 | 191065.1 | 148046.2 | 77.48471 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-13.csv | tat ca | 955929.2 | 1014190 | 93.90538 |
| C2013-13.csv | 25 | 955929.2 | 971770.8 | 98.34281 |
| C2013-13.csv | 1 -> 25 | 955929.2 | 575257.7 | 60.17786 |
| C2013-13.csv | 50 | 955929.2 | 1188967 | 75.62188 |
| C2013-13.csv | 1 -> 50 | 955929.2 | 784081.6 | 82.02298 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-03.csv | tat ca | 910728 | 954911.4 | 95.14856 |
| C2016-03.csv | 25 | 910728 | 885450.9 | 97.22451 |
| C2016-03.csv | 1 -> 25 | 910728 | 598833.3 | 65.75325 |
| C2016-03.csv | 50 | 910728 | 983744.1 | 91.98267 |
| C2016-03.csv | 1 -> 50 | 910728 | 795490.6 | 87.34667 |
| ----- | ------ | ------ | ------ | ------ |
| data\_020.csv | tat ca | 2512524 | 2054244 | 81.76018 |
| data\_020.csv | 25 | 2512524 | 1045079 | 41.59478 |
| data\_020.csv | 1 -> 25 | 2512524 | 947564.6 | 37.71365 |
| data\_020.csv | 50 | 2512524 | 2707610 | 92.23545 |
| data\_020.csv | 1 -> 50 | 2512524 | 1285312 | 51.15622 |
| ----- | ------ | ------ | ------ | ------ |
| data\_012.csv | tat ca | 494947.7 | 536858.7 | 91.53223 |
| data\_012.csv | 25 | 494947.7 | 361168.1 | 72.97095 |
| data\_012.csv | 1 -> 25 | 494947.7 | 246135.9 | 49.72968 |
| data\_012.csv | 50 | 494947.7 | 505722.7 | 97.82301 |
| data\_012.csv | 1 -> 50 | 494947.7 | 316824 | 64.01162 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-20.csv | tat ca | 11100 | 142277.6 | 0 |
| C2015-20.csv | 25 | 11100 | 142068.8 | 0 |
| C2015-20.csv | 1 -> 25 | 11100 | 141647.8 | 0 |
| C2015-20.csv | 50 | 11100 | 142663.1 | 0 |
| C2015-20.csv | 1 -> 50 | 11100 | 141858.3 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| data\_017.csv | tat ca | 1314585 | 958194.5 | 72.88953 |
| data\_017.csv | 25 | 1314585 | 488240.3 | 37.14027 |
| data\_017.csv | 1 -> 25 | 1314585 | 161154.5 | 12.25897 |
| data\_017.csv | 50 | 1314585 | 1021405 | 77.69792 |
| data\_017.csv | 1 -> 50 | 1314585 | 514705.1 | 39.15344 |
| ----- | ------ | ------ | ------ | ------ |
| C2014-02.csv | tat ca | 190266.5 | 270118.8 | 58.03132 |
| C2014-02.csv | 25 | 190266.5 | 211317.6 | 88.93599 |
| C2014-02.csv | 1 -> 25 | 190266.5 | 192375.5 | 98.89154 |
| C2014-02.csv | 50 | 190266.5 | 241239.5 | 73.20969 |
| C2014-02.csv | 1 -> 50 | 190266.5 | 209337.6 | 89.97664 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-20.csv | tat ca | 53783.28 | 160939.2 | 0 |
| C2016-20.csv | 25 | 53783.28 | 151597.6 | 0 |
| C2016-20.csv | 1 -> 25 | 53783.28 | 141052.5 | 0 |
| C2016-20.csv | 50 | 53783.28 | 165302.2 | 0 |
| C2016-20.csv | 1 -> 50 | 53783.28 | 146325 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-18.csv | tat ca | 60829.52 | 166034.6 | 0 |
| C2016-18.csv | 25 | 60829.52 | 155365.2 | 0 |
| C2016-18.csv | 1 -> 25 | 60829.52 | 142314.1 | 0 |
| C2016-18.csv | 50 | 60829.52 | 171680 | 0 |
| C2016-18.csv | 1 -> 50 | 60829.52 | 148839.6 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| Sheet21.csv | tat ca | 224789.7 | 335187.2 | 50.88852 |
| Sheet21.csv | 25 | 224789.7 | 261496.9 | 83.67042 |
| Sheet21.csv | 1 -> 25 | 224789.7 | 158565.5 | 70.53947 |
| Sheet21.csv | 50 | 224789.7 | 351295.1 | 43.72275 |
| Sheet21.csv | 1 -> 50 | 224789.7 | 210031.2 | 93.43453 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-09.csv | tat ca | 1696972 | 1557432 | 91.77716 |
| C2013-09.csv | 25 | 1696972 | 826293 | 48.69221 |
| C2013-09.csv | 1 -> 25 | 1696972 | 605169 | 35.6617 |
| C2013-09.csv | 50 | 1696972 | 1587723 | 93.56215 |
| C2013-09.csv | 1 -> 50 | 1696972 | 891965 | 52.56216 |
| ----- | ------ | ------ | ------ | ------ |
| Sheet19.csv | tat ca | 440940.9 | 475070.4 | 92.25983 |
| Sheet19.csv | 25 | 440940.9 | 342546.9 | 77.68546 |
| Sheet19.csv | 1 -> 25 | 440940.9 | 224381.2 | 50.88691 |
| Sheet19.csv | 50 | 440940.9 | 460332.1 | 95.60232 |
| Sheet19.csv | 1 -> 50 | 440940.9 | 320477.8 | 72.68044 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-10.csv | tat ca | 15218926 | 12532310 | 82.34687 |
| C2013-10.csv | 25 | 15218926 | 2170335 | 14.26076 |
| C2013-10.csv | 1 -> 25 | 15218926 | 1233404 | 8.104409 |
| C2013-10.csv | 50 | 15218926 | 9801687 | 64.40459 |
| C2013-10.csv | 1 -> 50 | 15218926 | 2371321 | 15.58139 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-08.csv | tat ca | 576624.1 | 637166.1 | 89.50061 |
| C2013-08.csv | 25 | 576624.1 | 430637 | 74.68245 |
| C2013-08.csv | 1 -> 25 | 576624.1 | 234025.9 | 40.58553 |
| C2013-08.csv | 50 | 576624.1 | 710698.3 | 76.7484 |
| C2013-08.csv | 1 -> 50 | 576624.1 | 413329.7 | 71.68097 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-06.csv | tat ca | 186107 | 307642.7 | 34.69581 |
| C2015-06.csv | 25 | 186107 | 228363.9 | 77.29431 |
| C2015-06.csv | 1 -> 25 | 186107 | 160661.8 | 86.32767 |
| C2015-06.csv | 50 | 186107 | 348497.8 | 12.74333 |
| C2015-06.csv | 1 -> 50 | 186107 | 226791 | 78.13945 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-02.csv | tat ca | 967988.8 | 820800.4 | 84.79441 |
| C2015-02.csv | 25 | 967988.8 | 646268.2 | 66.76401 |
| C2015-02.csv | 1 -> 25 | 967988.8 | 467164.8 | 48.26139 |
| C2015-02.csv | 50 | 967988.8 | 857072.2 | 88.54154 |
| C2015-02.csv | 1 -> 50 | 967988.8 | 583532.5 | 60.28299 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-10.csv | tat ca | 14466100 | 24656367 | 29.55761 |
| C2016-10.csv | 25 | 14466100 | 21483945 | 51.48765 |
| C2016-10.csv | 1 -> 25 | 14466100 | 25359402 | 24.69773 |
| C2016-10.csv | 50 | 14466100 | 25737429 | 22.08454 |
| C2016-10.csv | 1 -> 50 | 14466100 | 24364174 | 31.57745 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-27.csv | tat ca | 879701.1 | 704372 | 80.06947 |
| C2016-27.csv | 25 | 879701.1 | 586788.9 | 66.70322 |
| C2016-27.csv | 1 -> 25 | 879701.1 | 148977.5 | 16.93501 |
| C2016-27.csv | 50 | 879701.1 | 966779 | 90.10142 |
| C2016-27.csv | 1 -> 50 | 879701.1 | 367883.2 | 41.81912 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-17.csv | tat ca | 281070 | 325373.6 | 84.23751 |
| C2015-17.csv | 25 | 281070 | 154542.7 | 54.98371 |
| C2015-17.csv | 1 -> 25 | 281070 | 137157.9 | 48.7985 |
| C2015-17.csv | 50 | 281070 | 264609.2 | 94.14352 |
| C2015-17.csv | 1 -> 50 | 281070 | 145850.3 | 51.8911 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-06.csv | tat ca | 4331260 | 8036747 | 14.44785 |
| C2016-06.csv | 25 | 4331260 | 1058928 | 24.4485 |
| C2016-06.csv | 1 -> 25 | 4331260 | 416068.9 | 9.606185 |
| C2016-06.csv | 50 | 4331260 | 2527475 | 58.35427 |
| C2016-06.csv | 1 -> 50 | 4331260 | 1445675 | 33.37769 |
| ----- | ------ | ------ | ------ | ------ |
| C2013-15.csv | tat ca | 308343.8 | 437951.2 | 57.96658 |
| C2013-15.csv | 25 | 308343.8 | 228618.2 | 74.14393 |
| C2013-15.csv | 1 -> 25 | 308343.8 | 158295.1 | 51.33722 |
| C2013-15.csv | 50 | 308343.8 | 617041.7 | 0 |
| C2013-15.csv | 1 -> 50 | 308343.8 | 253419 | 82.18717 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-03.csv | tat ca | 1868796 | 1422066 | 76.0953 |
| C2015-03.csv | 25 | 1868796 | 737640.6 | 39.47143 |
| C2015-03.csv | 1 -> 25 | 1868796 | 205706.9 | 11.00745 |
| C2015-03.csv | 50 | 1868796 | 1041208 | 55.71544 |
| C2015-03.csv | 1 -> 50 | 1868796 | 523894.2 | 28.03378 |
| ----- | ------ | ------ | ------ | ------ |
| data\_18.csv | tat ca | 1425156 | 1536203 | 92.20806 |
| data\_18.csv | 25 | 1425156 | 1001511 | 70.2738 |
| data\_18.csv | 1 -> 25 | 1425156 | 418316.6 | 29.35234 |
| data\_18.csv | 50 | 1425156 | 1681055 | 82.04415 |
| data\_18.csv | 1 -> 50 | 1425156 | 839141.8 | 58.8807 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-23.csv | tat ca | 32805 | 149433.3 | 0 |
| C2015-23.csv | 25 | 32805 | 142539 | 0 |
| C2015-23.csv | 1 -> 25 | 32805 | 136628.7 | 0 |
| C2015-23.csv | 50 | 32805 | 156521.1 | 0 |
| C2015-23.csv | 1 -> 50 | 32805 | 139583.9 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2019-6.csv | tat ca | 1534060 | 1208739 | 78.79346 |
| C2019-6.csv | 25 | 1534060 | 284007.6 | 18.51346 |
| C2019-6.csv | 1 -> 25 | 1534060 | 143705.9 | 9.367681 |
| C2019-6.csv | 50 | 1534060 | 1038240 | 67.67921 |
| C2019-6.csv | 1 -> 50 | 1534060 | 392668.9 | 25.59671 |
| ----- | ------ | ------ | ------ | ------ |
| C2015-16.csv | tat ca | 64125 | 161560.4 | 0 |
| C2015-16.csv | 25 | 64125 | 137294.2 | 0 |
| C2015-16.csv | 1 -> 25 | 64125 | 136505 | 0 |
| C2015-16.csv | 50 | 64125 | 159749.8 | 0 |
| C2015-16.csv | 1 -> 50 | 64125 | 138926.5 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-25.csv | tat ca | 56748.33 | 162099.5 | 0 |
| C2016-25.csv | 25 | 56748.33 | 151375.3 | 0 |
| C2016-25.csv | 1 -> 25 | 56748.33 | 141758.5 | 0 |
| C2016-25.csv | 50 | 56748.33 | 165740.5 | 0 |
| C2016-25.csv | 1 -> 50 | 56748.33 | 146566.9 | 0 |
| ----- | ------ | ------ | ------ | ------ |
| C2016-32.csv | tat ca | 496991 | 527367.1 | 93.888 |
| C2016-32.csv | 25 | 496991 | 460066.9 | 92.57048 |
| C2016-32.csv | 1 -> 25 | 496991 | 236793.6 | 47.64546 |
| C2016-32.csv | 50 | 496991 | 539390.4 | 91.46878 |
| C2016-32.csv | 1 -> 50 | 496991 | 348430.3 | 70.10797 |
| ----- | ------ | ------ | ------ | ------ |