TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

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

-----🙞🙜🕮🙞🙜----- ****

**NGHIÊN CỨU CẢI TIẾN PHƯƠNG PHÁP SỬ DỤNG LSTM ĐỂ DỰ ĐOÁN MỨC ĐỘ HOÀN THÀNH DỰ ÁN PHẦN MỀM**

|  |  |  |
| --- | --- | --- |
| Giáo viên hướng dẫn : ***PGS. TS. Huỳnh Quyết Thắng*** | | |
| Thực hiện:  Lớp: | **Đỗ Tiến Đạt**  **Nguyễn Bá Hùng**  **Nguyễn Viết Luyện**  **KSTN CNTT K60**  **Hà Nội 12/2019** |  |

**LỜI CAM ĐOAN**

Tập thể các sinh viên:

1. Đỗ Tiến Đạt, lớp KSTN CNTT K60
2. Nguyễn Bá Hùng, lớp KSTN CNTT K60
3. Nguyễn Viết Luyện, lớp KSTN CNTT K60

Xin cam đoan đây là công trình nghiên cứu, cải tiến của tập thể nhóm sinh viên KSTN CNTT K60 chúng em dựa trên đề tài của các sinh viên lớp KSTN CNTT K59 dưới sự hướng dẫn của PGS.TS Huỳnh Quyết Thắng, Bộ môn Công nghệ phần mềm – Viện Công nghệ thông tin và Truyền thông – Trường Đại học Bách Khoa Hà Nội, và Nghiên cứu sinh, Thạc sỹ Lê Thế Anh.

Các tài liệu tham khảo trong nghiên cứu được liệt kê tại phần tài liệu tham khảo ở cuối báo cáo, cho nội dung trích dẫn đã ghi rõ nguồn gốc.

Các số liệu, kết quả nêu trong báo cáo là trung thực, rõ ràng.

Hà Nội, ngày … tháng … năm 2019

**Nhóm thực hiện nghiên cứu**

**LỜI CẢM ƠN**

Lời đầu tiên chúng tôi xin dành lời cảm ơn chân thành và sâu sắc tới thầy PGS.TS Huỳnh Quyết Thắng, bộ môn Công nghệ phần mềm – Viện Công nghệ thông tin và Truyền thông – Trường Đại học Bách Khoa Hà Nội.

Trong quá trình thực hiện nghiên cứu, Thầy đã hướng dẫn và chỉ bảo chúng tôi ân cần, cung cấp cho chúng tôi nhiều tài liệu quan trọng là nền tảng để tôi tìm tòi nghiên cứu sâu hơn về đề tài “Nghiên cứu cải tiến phương pháp LSTM để dự đoán mức độ hoàn thành dự án phần mềm”. Mỗi lời góp ý của Thầy giúp chúng tôi định hướng tốt hơn, đúng đắn hơn cho đề tài mình đã chọn.

Chúng tôi cũng xin gửi lời cảm ơn tới Thạc sỹ Lê Thế Anh đã tận tình giúp đỡ, và cung cấp nhiều kiến thúc và tài liệu bổ ích và cần thiết. Để từ đó, chúng tôi đủ khả năng và hiểu biết để thực hiện và hoàn thành nghiên cứu này một cách tốt nhất.

Lời tiếp theo chúng tôi xin cảm ơn các anh lớp KSTN CNTT K59 với đề tài “Ứng dụng mạng nơ ron nhân tạo để khai phá dữ liệu EVM nhằm nâng cao chất lượng dự đoán mức độ hoàn thành phần mềm” là nền tảng cơ bản để tiếp tục phát triển nghiên cứu cải tiến này. Cuối cùng chúng tôi xin dành lời cảm ơn tới bạn bè, các thành viên trong lớp KSTN CNTT K60 đã quan tâm, chia sẻ, ủng hộ để chúng tôi có động lực học tập và hoàn thành nghiên cứu này.

Mục lục

[CHƯƠNG 1: CƠ SỞ LÝ THUYẾT 5](#_Toc29498079)

[1.1 Kỹ thuật quản trị giá trị thu được trong quản lý dự án 5](#_Toc29498080)

[1.1.1 Tổng quan 5](#_Toc29498081)

[1.1.2 Các đặc trưng 5](#_Toc29498082)

[1.2 Phương pháp LSTM trong áp dụng kỹ thuật EVM 6](#_Toc29498083)

[1.3 Sơ đồ kiến trúc mạng LSTM áp dụng cho bài toán EVM 9](#_Toc29498084)

[CHƯƠNG 2: XÂY DỰNG MẠNG NƠ-RON NHÂN TẠO LSTM ĐỂ ƯỚC LƯỢNG THỜI GIAN VÀ CHI PHÍ DỰ ÁN 10](#_Toc29498085)

[2.1. Xây dựng tập dữ liệu 10](#_Toc29498086)

[2.1.1: BỘ DỮ LIỆU K59: 10](#_Toc29498087)

[2.1.2: BỘ DỮ LIỆU K60 12](#_Toc29498088)

[2.2. Xây dựng mô hình mạng nơ-ron nhân tạo 13](#_Toc29498089)

[CHƯƠNG 3: CÀI ĐẶT, KẾT QUẢ THỬ NGHIỆM 15](#_Toc29498090)

[3.1 Xây dựng môi trường cài đặt và thử nghiệm: 15](#_Toc29498091)

[3.2 Kết quả thực nghiệm trên bộ dữ liệu K59 15](#_Toc29498092)

[3.2.1 Mô hình AC, XT dự đoán TOTAL\_AC; So sánh với kết quả nhóm các anh K59 15](#_Toc29498093)

[3.2.2: Mô hình AT, XT dự đoán TOTAL\_AT: 20](#_Toc29498094)

[3.3. Kết quả thực nghiệm trên bộ dữ liệu K60 22](#_Toc29498095)

[3.3.1. Thực hiện mô hình XT, AC đoán TOTAL\_AC (Mô hình K60) 22](#_Toc29498096)

[3.3.2. Thực hiện mô hình XT, AT đoán TOTAL\_AT (mô hình K60) 23](#_Toc29498097)

[3.4 Đánh giá mô hình: 24](#_Toc29498098)

[KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN 25](#_Toc29498099)

# CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

Chương này trình bày các vấn đề lý thuyết sau:

* Các kỹ thuật quản trị giá trị thu được trong quản lý dự án
* Tìm hiểu phương pháp LSTM trong áp dụng kỹ thuật EVM (sử dụng đồ án môn học của nhóm sinh viên 2018-2019 về sử dụng LSTM và EVM trong đánh giá chi phí của dự án
* Mô tả phương pháp LSTM dưới dạng thuật toán

# 1.1 Kỹ thuật quản trị giá trị thu được trong quản lý dự án

## 1.1.1 Tổng quan

Quản lý giá trị thu được (tiếng Anh là Earned Value Management, viết tắt EVM) là một kỹ thuật quản lý dự án để đo lường sự tiến triển (tiến trình thực hiện) của dự án một cách khách quan.

Quản lý giá trị thu được có khả năng kết hợp các phép đo về phạm vi, tiến độ và chi phí trong một hệ thống tích hợp duy nhất. Khi áp dụng đúng, việc quản lý giá trị thu được sẽ cung cấp một cảnh báo sớm về những vấn đề thực thi dự án. Ngoài ra, thuật quản lý giá trị thu được hứa hẹn cải thiện việc xác định rõ phạm vi dự án, ngăn chặn sự mất kiểm soát phạm vi dự án (Scope creep), truyền đạt về tiến trình mục tiêu tới các bên liên quan và giữ cho nhóm dự án tập trung vào việc đạt được tiến bộ.

Quản lý giá trị thu được là kỹ thuật kiểm soát sự thực hiện của dự án, kết hợp được cả việc kiểm soát chi phí lẫn kiểm soát tiến độ dự án. Quản lý giá trị thu được chính là kỹ thuật kiểm soát chi phí của dự án, nhưng gắn liền hữu cơ với việc kiểm soát tiến độ thực hiện.

## 1.1.2 Các đặc trưng

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

* Một bản kế hoạch dự án (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 và kế hoạch về thời gian thực hiện.
* 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 PV (Planned Value) hoặc Dự toán ngân quỹ chi phí cho công việc theo tiến độ BCWS (Budgeted Cost of Work Scheduled),
* 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 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 BCWP (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 (tức là vượt ngân quỹ, đồng nghĩa bị lỗ) hoặc (dưới ngân quỹ, có lãi) trong vòng ngân quỹ) và tiến độ thực hiện SPI (chậm tiến độ 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.

# 1.2 Phương pháp LSTM trong áp dụng kỹ thuật EVM

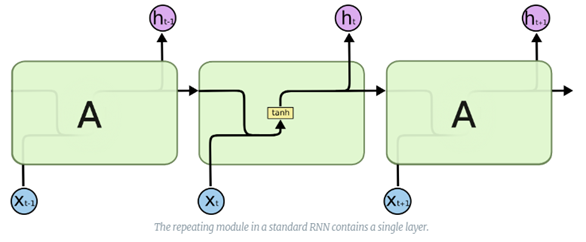
Mạng nơ-ron nhân tạo hay thường gọi ngắn gọn mạng nơ-ron là một mô hình toán học hay mô hình tính toán được xây dựng dựa trên các mạng nơ-ron sinh học. Nó gồm có một nhóm các nơ-ron nhân tạo (nút) nối với nhau, và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút. Trong nhiều trường hợp, mạng nơ-ron nhân tạo là một hệ thống thích ứng (adaptive system) tự thay đổi cấu trúc của mình dựa trên các thông tin bên ngoài hay bên trong chảy qua mạng trong quá trình học.

Trong thực tế sử dụng, nhiều mạng nơ-ron là các công cụ mô hình hóa dữ liệu thống kê phi tuyến. Chúng có thể được dùng để mô hình hóa các mối quan hệ phức tạp giữa dữ liệu vào và kết quả hoặc để tìm kiếm các dạng/mẫu trong dữ liệu.

Mô hình đề xuất dựa trên mô hình mạng neural LSTM, đây là một dạng đặc biệt của RNN (Recurrent Neural Network - Mạng neural hồi quy). LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997), và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau.

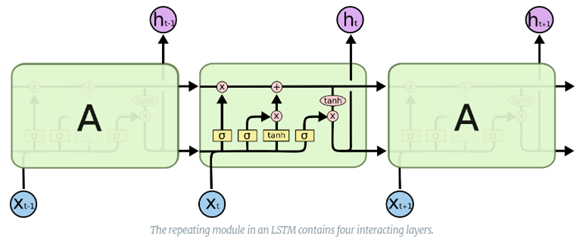
Mạng LSTM được sử dụng khá phổ biến hiện nay. Về cơ bản mô hình của LSTM không khác mô hình truyền thống của RNN, nhưng chúng sử dụng hàm tính toán khác ở các trạng thái ẩn.Mạng LSTM có khả năng học được các phụ thuộc xa (long-term dependency). Việc nhớ thông tin trong suốt thời gian dài là đặc tính mặc định của chúng, chứ ta không cần phải huấn luyện nó để có thể nhớ được. Tức là nó có thể ghi nhớ được mà không cần can thiệp từ bên ngoài.

Mọi mạng hồi quy đều có dạng là một chuỗi các mô-đun lặp đi lặp lại của mạng nơ-ron. Với mạng RNN chuẩn, các mô-đun này có cấu trúc rất đơn giản, thường là một tầng tanh.



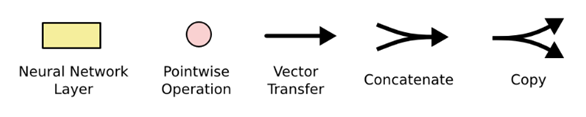
Hình 1. Mô hình cấu trúc mạng RNN chuẩn

LSTM cũng có kiến trúc dạng chuỗi như vậy, nhưng các mô-đun trong nó có cấu trúc khác với mạng RNN chuẩn. Thay vì chỉ có một tầng mạng nơ-ron, chúng có tới 4 tầng tương tác với nhau một cách rất đặc biệt.



*Hình 2. Mô hình cấu trúc mạng LSTM*

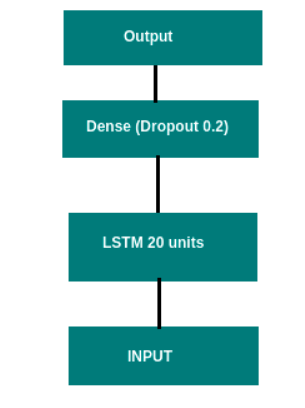
Trong đó:



*Hình 3. Ký hiệu sử dụng trong mô hình mạng LSTM*

Trong sơ đồ ta thấy mỗi một đường mang một vector từ đầu ra của một nút tới đầu vào của một nút khác. Các hình trong màu hồng biểu diễn các phép toán, ví dụ như phép cộng vector, còn các ô màu vàng được sử dụng để học trong từng mạng nơ-ron. Các đường hợp nhau kí hiệu việc kết hợp, còn các đường rẽ nhánh ám chỉ nội dung của nó được sao chép và chuyển tới các nơi khác nhau.

# 1.3 Sơ đồ kiến trúc mạng LSTM áp dụng cho bài toán EVM

****

Hình 4. Sơ đồ cấu trúc mạng LSTM cho bài toán EVM

Gồm 1 input đầu vào và 3 hidden layer:

* 1 input đầu vào: Có thể là AT, XT, AC
* 1 layer cho LSTM với 20 units, như đã trình bày phía trên của báo cáo.
* Khi qua LSTM 20 units thì nó trả về shape (NONE, 20), ta cho qua một tầng hidden layer nữa cho dropout 0.2
* Khi qua tầng này, ta cho qua một tầng dense cuối cùng, và đó là output đầu ra của chúng ta.

# CHƯƠNG 2: XÂY DỰNG MẠNG NƠ-RON NHÂN TẠO LSTM ĐỂ ƯỚC LƯỢNG THỜI GIAN VÀ CHI PHÍ DỰ ÁN

Trong chương này sẽ trình bày các vấn đề:

* Xây dựng tập dữ liệu dùng để huấn luyện
* Xây dựng mô hình mạng nơ-ron nhân tạo LSTM

# 2.1. Xây dựng tập dữ liệu

Để thực hiện huấn luyện và kiểm thử độ chính xác của mô hình, nhóm thực hiện trên 2 bộ dữ liệu: Gồm bộ dữ liệu của các anh K59 đã làm năm trước, nhóm gọi là **BỘ DỮ LIỆU K59,** và bộ dữ liệu gồm 22 bộ dự án công nghệ phần mềm từ nước ngoài (được gửi theo tệp đính kèm) do nhóm tập hợp và xử lý từ trang web <http://www.projectmanagement.ugent.be/research/data/realdata> thời gian các dự án dao động từ 4 đến 30 tháng, nhóm gọi là **BỘ DỮ LIỆU K60.**

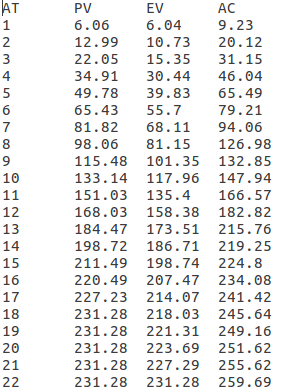
## 2.1.1: BỘ DỮ LIỆU K59:

Bộ dữ liệu này được đính kèm trong thư mục: ***code/k59/data/***

Bộ này được thành 2 bộ:

* Bộ train: Gồm 7 bộ: dataA.txt->dataG.txt, nằm trong thư mục: ***code/k59/data/train***
* Bộ test: Bộ dataH.txt, nằm trong thư mujc: ***code/k59/data/test***

***Dữ liệu thô của dự án:***

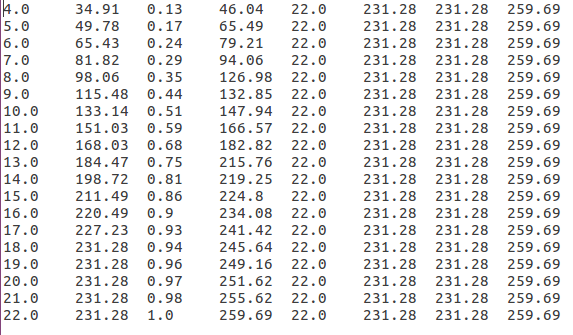


Bảng 1. Dữ liệu thô của dự án trong dự án KSTN CNTT K59

Trong đó:

* Cột 1 là AT: thời gian thực hiện dự án (đo bằng tháng)
* Cột 2 là PV: chi phí được lên kế hoạch (đo bằng tiền)
* Cột 3 là XT: tiến độ thực hiện dự án (XT= EVcurrent/TOTAL\_EV)
* Cột 4 là AC: chi phi bỏ ra tính đến hiện tại (đo bằng tiền)

***Dữ liệu sau khi xử lý:***



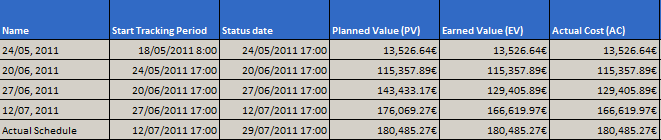
Bảng 2. Dữ sau xử lý của dự án trong dự án KSTN CNTT K59

Trong đó:

* Cột 1 là AT: thời gian thực hiện dự án (đo bằng tháng)
* Cột 2 là PV: chi phí được lên kế hoạch (đo bằng tiền)
* Cột 3 là XT: tiến độ thực hiện dự án (XT= EVcurrent/TOTAL\_EV)
* Cột 4 là AC: chi phi bỏ ra tính đến hiện tại (đo bằng tiền)
* Cột 5 là TOTAL\_AT: tổng số thời gian thực hiện dự án
* Cột 6 là TOTAL\_PV: tổng chi phí hoàn thành dự án được lên kế hoạch
* Cột 7 là TOTAL\_EV: tổng chi phí ước lượng đạt được của dự án
* Cột 8 là TOTAL\_AV: tổng chi phí bỏ ra để hoàn thành dự án

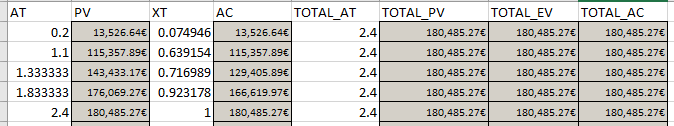
## 2.1.2: BỘ DỮ LIỆU K60

***Dữ liệu thô của dự án:***



Bảng 3. Dữ liệu thô của dự án trong dự án KSTN CNTT K60

***Dữ liệu sau khi được xử lý:***



Bảng 4. Dữ liệu sau xử lý của dự án trong dự án KSTN CNTT K60

Trong đó:

* AT: thời gian thực hiện dự án (đo bằng tháng)
* 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

# 2.2. Xây dựng mô hình mạng nơ-ron nhân tạo

Xây dựng 3 mô hình LSTM khác nhau:

**Mô hình AT, XT dự đoán TOTAL\_AT; AC, XT dự đoán TOTAL\_AC:**

* Giống mô hình ở trên đã trình bày
* Đầu vào: Có shape (None, 3, 2) tương ứng như sau:
  + None ở đây thể hiện là batch\_size
  + 3 ở đây là time\_steps, ở đây của mình có nghĩa là dùng dữ liệu 3 công việc để đoán tổng cộng cho dự án. (TOTAL\_AT, TOTAL\_AC)
  + 2 ở đây là 2 feature, ở trong bài toán thì nó là AT,XT hoặc AC,XT
* Tiếp theo là một Hidden Layer LSTM:
  + Gồm có 20 units.
  + Hàm activation là RELU
  + Khi qua hàm này, thì nó sẽ trả về một shape: (None, 20)
* Tiếp theo sẽ là một hidden layer dùng để Dropout:
  + Ở đây mình cho dropout: 0.2
* Sau khi qua dropout, thì mình sẽ có 1 tầng Dense:
  + Sau khi qua tầng này thì nó đóng vai trò là đầu ra luôn
  + Đầu ra có shape: (None, 1)
  + 1 ở đây là 1 đầu ra tương ứng với TOTAL\_AC hoặc TOTAL\_AT

**Mô hình AT, XT, AC dự đoán TOTAL\_AT, TOTAL\_AC theo mô hình many to one**

* Giống mô hình ở trên đã trình bày
* Đầu vào: Có shape (None, 3, 3) tương ứng như sau:
  + None ở đây thể hiện là batch\_size
  + 3 ở đây là time\_steps, ở đây của mình có nghĩa là dùng dữ liệu 3 công việc để đoán tổng cộng cho dự án. (TOTAL\_AT, TOTAL\_AC)
  + 3 cuối cùng ở đây là 3 feature, ở trong bài toán thì nó là AT,AC,XT
* Tiếp theo là một Hidden Layer LSTM:
  + Gồm có 20 units.
  + Hàm activation là RELU
  + Khi qua hàm này, thì nó sẽ trả về một shape: (None, 20)
* Tiếp theo sẽ là một hidden layer dùng để Dropout:
  + Ở đây mình cho dropout: 0.2
* Sau khi qua dropout, thì mình sẽ có 1 tầng Dense:
  + Sau khi qua tầng này thì nó đóng vai trò là đầu ra luôn
  + Đầu ra có shape: (None, 2)

1 ở đây là 1 đầu ra tương ứng với TOTAL\_AC và TOTAL\_AT.

# CHƯƠNG 3: CÀI ĐẶT, KẾT QUẢ THỬ NGHIỆM

Chương này trình bày:

* Xây dựng môi trường cài đặt và thử nghiệm
* Tính toán chi phi và thời gian hoàn thành dự án theo các mô hình
* Đánh giá, so sánh các mô hình

# 3.1 Xây dựng môi trường cài đặt và thử nghiệm:

**Phần cứng:**

Google Colab

**Phần mềm:**

Ngôn ngữ lập trình: Python 3

Thư viện lập trình: Thư viện học máy: Tensorflow, keras;

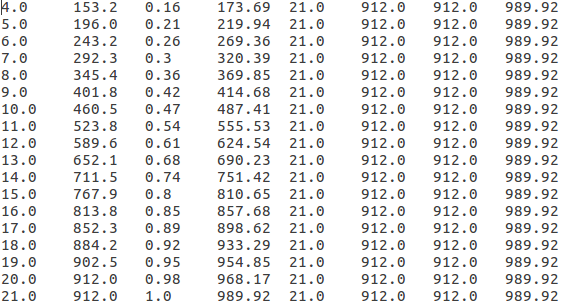
# 3.2 Kết quả thực nghiệm trên bộ dữ liệu K59

Về bộ dữ liệu này, nhóm đã lấy của các anh K59, để thực hiện các mô hình bên dưới. Đồng thời, nhóm dựa vào dữ liệu này, để có thể so sánh độ chính xác của mô hình (AC, XT dự đoán TOTAL\_AC) của nhóm so với các anh K59 (do các anh K59 chỉ làm được mỗi mô hình kiểu này)

## 3.2.1 Mô hình AC, XT dự đoán TOTAL\_AC; So sánh với kết quả nhóm các anh K59

Sau khi huấn luyện các mô hình từ các dữ liệu Train thì mình tiến hành kiểm nghiệm xem mô hình nhóm em với nhóm k59 theo bộ test: dataH.txt.

Cụ thể thì dataH.txt như sau:



Bảng 5. Bộ test dữ liệu dataH.txt

Do cách xử lý dữ liệu để đưa vào mô hình của các anh K59, và nhóm em là khác nhau, nên để đánh giá chính xác giữa 2 mô hình, em sẽ dựa trên cùng chung số liệu (xt,ac), ví dụ, với bộ dữ liệu dataH thì: (Nên nhớ em đã train theo kiểu dùng 3 cặp (xt,ac) để có thể đoán ra TOTAL\_AC của dự án).

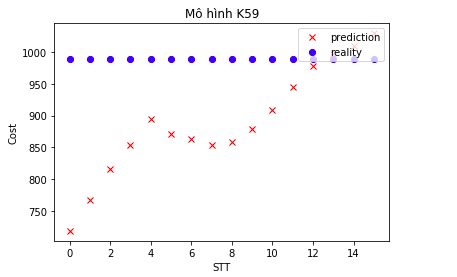
**Input:** Gồm 3 cặp (xt, ac) là [(0.16, 173.69), (0.21, 219.94), (0.26, 269.36)]

**Thì :** Mô hình K59 đoán TOTAL\_AC là a; Mô hình K60 đoán TOTAL\_AC là b thì khi đó mình sẽ tính % tính đúng của từng mô hình bằng cách: Mô hình K59 tính đúng (a/989.92)\*100% với a <=989.92 hoặc 1-(a/989.92-1)\*100% với a >989.92 ; Mô hình K60 tính đúng là (b/989.92)\*100% với b <=989.92 hoặc 1-(b/989.92-1)\*100% với b >989.92. (**Lưu ý** TOTAL\_AC của một dự án chỉ có một giá trị duy nhất).

**Cuối cùng:** Sau khi đã có độ chính xác của từng nhóm đầu vào rồi em sẽ tính trung bình chúng xem mô hình nào tốt hơn.

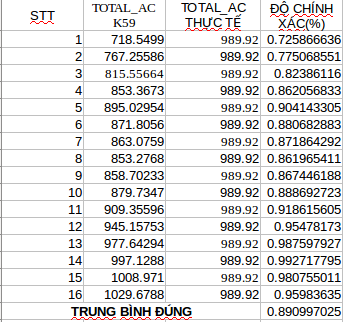
* ***Mô hình XT, AC dự đoán TOTAL\_AC của các anh K59:***

**Kết quả:**

****

Hình 10. Kết quả mô hình dự đoán KSTN CNTT K59

**Số liệu:** (Lưu ý: STT ở trên hình không phải là thời gian tính theo từng tháng, ví dụ số 2 ở cột x không phải là 2 tháng mà là nhóm thứ 2 gồm 3 cặp (xt, ac) liên tiếp để đoán ra TOTAL\_AC)

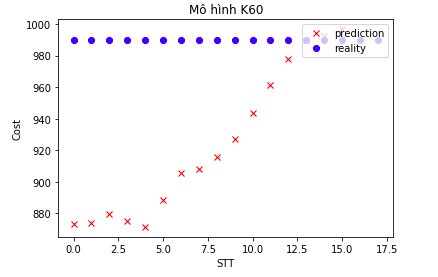


Bảng 6. Độ chính xác mô hình dự đoán TOTAL\_AC theo mô hình KSTN K59

Trung bình chính xác là 89.1%

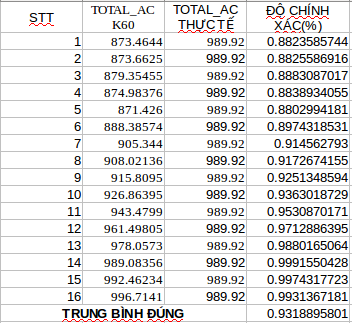
* ***Mô hình XT, AC dự đoán TOTAL\_AC của nhóm:***

**Kết quả**

****

Hình 11. Kết quả mô hình dự đoán KSTN CNTT K60

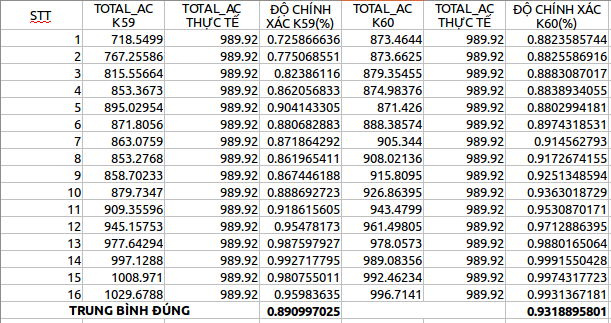
**Số liệu:** (Lưu ý: STT ở trên hình không phải là thời gian tính theo từng tháng, ví dụ số 2 ở cột x không phải là 2 tháng mà là nhóm thứ 2 gồm 3 cặp (xt,ac) liên tiếp để đoán ra TOTAL\_AC)



Bảng 7. Độ chính xác mô hình dự đoán TOTAL\_AC theo mô hình KSTN CNTT K60

Trung bình chính xác là 93.2 %

* **SO SÁNH MÔ HÌNH K59 VỚI K60:**

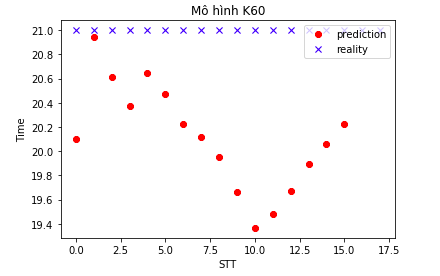
****

Bảng 8. Bảng so sánh độ chính xác mô hình dự đoán TOTAL\_AC theo mô hình KSTN CNTT K59 & KSTN CNTT K60

Dựa vào số liệu chúng ta có thể thấy **mô hình K60 với độ chính xác 93.2%** tốt hơn so với **mô hình của K59 với độ chính xác là 89.1%**.

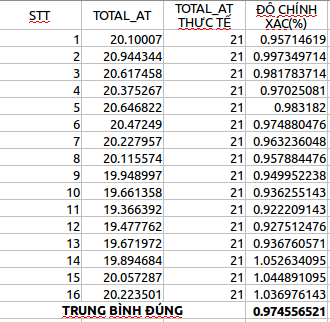
## 3.2.2: Mô hình AT, XT dự đoán TOTAL\_AT:

**Kết quả**

****

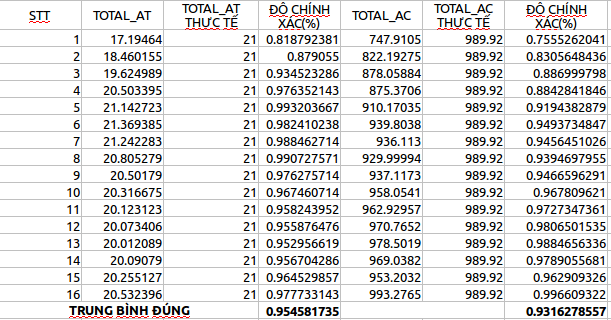
Hình 12 Kết quả mô hình dự đoán TOTAL\_AT KSTN CNTT K60

**Số liệu**

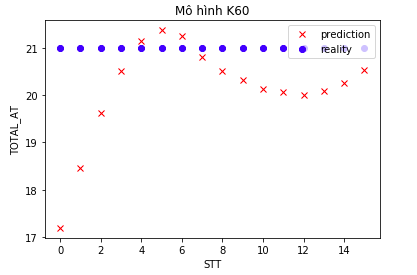
****

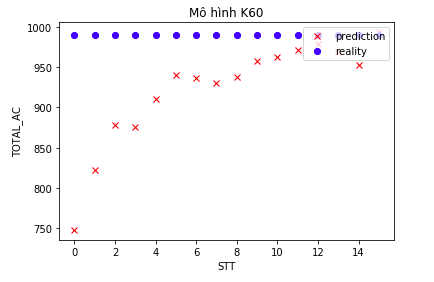
Bảng 9. Độ chính xác của mô hình dự đoán TOTAL\_AT

## 3.2.3: Mô hình AT, XT, AC dự đoán TOTAL\_AT, TOTAL\_AC :

****

Bảng 10. Độ chính xác mô hình dự đoán đồng thời TOAL\_AT và TOTAL\_AC KSTN CNTT K60

****

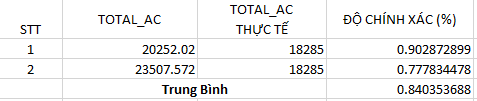
****

Hình 13. Kết quả dự đoán mô hình output đồng thời total\_AC và total\_AT

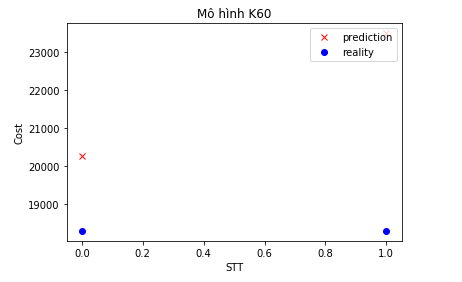
**3.3. Kết quả thực nghiệm trên bộ dữ liệu K60**

Về bộ dữ liệu này, nhóm đã tự lấy trên trang web nêu ở trên, để thực hiện các mô hình bên dưới. Đồng thời, nhóm dựa vào dữ liệu này, để có thể so sánh độ chính xác của mô hình (AC, XT dự đoán TOTAL\_AC) của nhóm so với các anh K59 (do các anh K59 chỉ làm được mỗi mô hình kiểu này)

**3.3.1. Thực hiện mô hình XT, AC đoán TOTAL\_AC (Mô hình K60)**

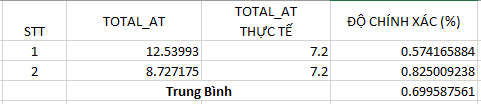
****

Bảng 11. Độ chính xác của mô hình dự đoán TOTAL\_AC

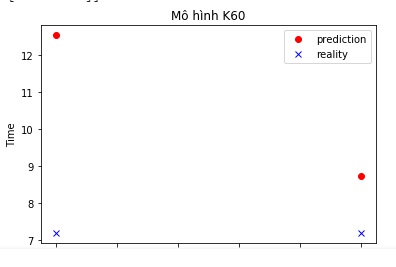
****

Hình 14: Kết quả dự đoán TOTOAL\_AC mô hình K60

**3.3.2. Thực hiện mô hình XT, AT đoán TOTAL\_AT (mô hình K60)**

****

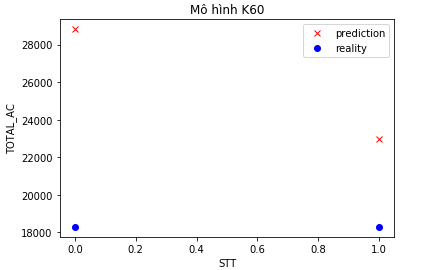
Bảng 12. Độ chính xác của mô hình dự đoán TOTAL\_AT

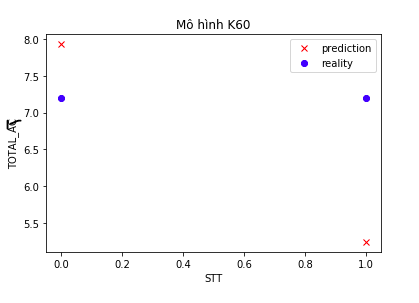
****

Hình 15: Kết quả dự đoán TOTAL\_AT mô hình K60

**3.3.3. Thực hiện mô hình XT, AT, AC đoán TOTAL\_AT (mô hình K60)**

****

****



Hình 16. Kết quả mô hình XT, AT, AC đoán TOTAL\_AT (mô hình K60)

**3.3.4. Thực hiện mô hình XT, AT, AC đoán TOTAL\_AT (mô hình K59)**



Bảng 13. Kết quả mô hình XT, AT, AC đoán TOTAL\_AT (mô hình K59)

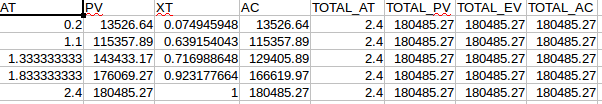
**Mô hình này cho kết quả rất kém.**

# 3.4 Đánh giá mô hình:

Sau khi so sánh mô hình của nhóm so với các anh K59 trên cùng bộ dữ liệu mà các anh đã có thì nhóm thấy kết quả của nhóm thu được tốt hơn hẳn so với các anh K59. Điều này cho thấy về thuật toán nhóm đã xử lý tốt hơn so với mô hình cũ, ở cả 2 loại mô hình.

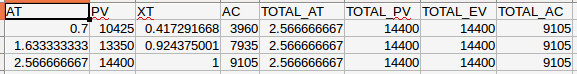
Còn mô hình của nhóm với dữ liệu mới này, thì kết quả chưa thực sự tốt, bởi thuật toán LSTM quan trọng dữ liệu đầu vào, nhưng dữ liệu hiện tại của nhóm có một số nhược điểm sau:

* Dữ liệu thu được từ các dự án khác nhau, mỗi dự án lại có đặc điểm khác nhau về tính chất như quy mô, thời gian, năng suất khác nhau. Ví dụ:
  + Dự án A: Trong 1 Tháng với tiến độ 50% thì Actual Cost là 8 triệu, nhưng có dự án trong 1 tháng với tiến độ 50% thì actual cost lại 100 triệu. Với số liệu như thế này thì LSTM khó có thể học được, chỉ giải quyết được những bài toán này khi dữ liệu rất nhiều.
  + Hiện tại nhóm sẽ lấy cụ thể một số dự án mà nhóm đã thu thập được, để chứng minh vấn đề của nhóm vẫn là dữ liệu chứ không phải là thuật toán: Ví dụ trong data\_1.csv



Bảng 14. data\_1.csv

và data\_3.csv



Bảng 15. Data\_3.csv

* Ta thấy với data\_1 thì 0.2 tháng với tiến độ 0.0749 thì đã có AC là 13526.64 trong khi data\_3 với tiến độ 0.417 thì AC chỉ là 3960, kể cả 2.56 tháng với tiến độ 1 rồi AC cũng chỉ là 9105.
* Vì dữ liệu quá ít với lại những số liệu từ những dự án thế này sẽ gây nhiễu và làm cho LSTM không thể học chính xác được.
* Dữ liệu thu được còn quá ít, dù có tính chất khác nhau của các dự án.

# KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN

**Kết luận**

Nghiên cứu với đề tài “Nghiên cứu cải tiến phương pháp sử dụng LSTM để dự đoán mức độ hoàn thành dự án phần mềm” cơ bản hoàn thành và thực hiện các nhiệm vụ chính đã đặt ra. Nhiệm vụ đầu tiên là tìm hiểu khái quát về các các kỹ thuật quản lý giá trị thu được, phương pháp LSTM áp dụng kỹ thuật EVM. Nhiệm vụ thứ hai là xác định dữ liệu đầu vào cho bài toán và phân tích dữ liệu, chuyển đổi cấu trúc dữ liệu về dạng cần cho quá trình huấn luyện. Từ đó đề xuất xây dựng mô hình mạng nơ-ron LSTM giải quyết bài toán đã đặt ra, dự đoán chi phí về tổng thời gian thực hiện dự án và tổng chi phí bỏ ra để hoàn thành dự án. Cuối cùng cài đặt chương trình dựa trên ngôn ngữ Python dựa trên mô hình đề xuất. Đánh giá mô hình đề xuất, so sánh kết quả thực nghiệm giữa các mô hình từ đó tìm ra mô hình tốt nhất.

Nghiên cứu đã đạt được một số kết quả. Đầu tiên là chúng ta đã tổng hợp được các cơ sở lý thuyết. Phần lý thuyết trình bày chi tiết, dễ hiểu, xúc tích về EVM, các phương pháp phân tích tập trung tìm hiểu chi tiết lý thuyết các phương pháp áp dụng để dự báo chi phí và thời gian hoàn thành dự án, đặc biệt về ứng dụng mạng nơ-ron nhân tạo. Kết quả thực nghiệm cho thấy khá khả quan, từ đó khẳng định tính đúng đắn của phương pháp đề xuất, khẳng định dự đoán chi phí và thời gian là khả thi khi sử dụng mạng nơ-ron.

Trong quá trình thực hiện nghiên cứu, chúng tôi đã gặp một số khó khăn. Khó khăn đầu tiên là việc tìm kiếm dữ liệu để áp dụng trong quá trình huấn luyện. Do dữ liệu thực tế lấy được là không tương đồng, khoảng cách chi phí và thời gian từ thấp nhất đến cao nhất là không giống nhau. Điều này gây khó khăn khi đề xuất mô hình, huấn luyện trọng số phù hợp với nhiều bộ dữ liệu nhất. Khó khăn thứ hai là việc xây dựng mô hình LSTM dự đoán. Do đây là một mô hình còn khá mới mẻ với nhóm, để có thể sử dụng thì nhóm phải tìm hiểu nghiên cứu các tài liệu về LSTM, từ đó kết hợp với phương pháp EVM để đưa ra các mô hình hợp lý nhất. Khó khăn thứ ba là việc huấn luyện cần yêu cầu khối lượng xử lý lớn, thời gian huấn luyện lâu mà máy tính cá nhân không thể đáp ứng. Để giải quyết nhóm phải tìm hiểu sử dụng các công cụ hỗ trợ như Google Colab để huấn luyện mô hình, đưa ra kết quả các mô hình đề xuất.

**Định hướng phát triển**

Nghiên cứu đã đưa ra một số mô hình mới để dự đoán tổng chi phí và thời gian hoàn thành dự án phần mềm. Mỗi mô hình đều được cải tiến, qua đó dự đoán tốt hơn yêu cầu đặt ra. Trong tương lai chúng tôi sẽ thử nghiệm các mô hình mới, được cho là có thể phù hợp, so sánh kết quả mô hình cũ, từ đó tìm ra những mô hình tốt hơn nữa.

Ngoài ra, dữ liễu đóng vai trò quan trọng để huấn luyện ra mô hình tốt. Chúng ta có thể tìm những bộ dữ liệu mới, có độ tương đồng cao hơn, bổ sung thêm vào tập dữ liệu hiện có. Qua đó càng làm mô hình có thể chính xác hơn.

# TÀI LIỆU THAM KHẢO

[1] Đồ án môn học của nhóm sinh viên năm 2018-2019 về sử dụng LSTM và EVM trong đánh giá chi phí của dự án (Báo cáo, mã nguồn Python và dữ liệu)

[2] Jordy Batselier and Mario Vanhoucke (2015), Emprical Evaluation of Earned Value Management Forecasting Accurary for Time and Cost

[3] Dữ liệu các project thực tế các project công nghệ phần mềm:

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

[4] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. Neural computation, 1997, 9.8: 1735-1780

[5] Long short-term memory Wikipedia

<https://en.wikipedia.org/wiki/Long_short-term_memory>

[6] LSTM là gì?

<https://dominhhai.github.io/vi/2017/10/what-is-lstm/>