

Acad. Year

2020-21

Project No.

A123

Thuộc bài tập lớn Phát hiện gian lận tín dụng

# XGBoost



- Nguyễn Hải Long 18020032
- Nguyễn Việt Long 18020793
- Trịnh Tuấn Tú 18021352

Đại học Công Nghệ - Đại học Quốc Gia Hà Nội

XGBoost

**Nguyễn Hải Long 18020032**

**Nguyễn Việt Long 18020793**

**Trịnh Tuấn Tú 18021352**

Đại học Công Nghệ & khoa Khoa học Máy Tính

Báo cáo kết quả nghiên cứu bài tập lớn môn Học Máy

Bài toán phát hiện gian lận tín dụng

Năm học 2020/21

# Mục lục

<b>1</b>	<b>Giới thiệu</b>	<b>2</b>
1.1	Thuật toán Gradient Tree Boosting . . . . .	2
1.2	XGBoost - một cải tiến của Gradient Tree Boosting . . . . .	2
<b>2</b>	<b>Thuật toán Gradient Tree Boosting</b>	<b>3</b>
2.1	Regularized hàm mục tiêu . . . . .	3

# **1 Giới thiệu**

## **1.1 Thuật toán Gradient Tree Boosting**

## **1.2 XGBoost - một cải tiến của Gradient Tree Boosting**

## 2 Thuật toán Gradient Tree Boosting

### 2.1 Regularized hàm mục tiêu

Cho một tập dữ liệu có  $n$  mẫu, và mỗi mẫu có  $m$  đặc trưng  $\mathcal{D} = \{(x_i, y_i)\} (|\mathcal{D}| = n, x_i \in \mathbb{R}^m), y_i \in \mathbb{R}$ . Một hình Tree-Boosting là việc đưa  $x_i$  qua  $K$  cây quyết định rồi sau đó sử dụng một hàm cộng tất cả các kết quả đầu ra của các cây đó làm kết quả. Công thức thường hay được sử dụng được mô tả như công thức bên dưới.

$$\mathbb{R} \tag{1}$$