



## **Data Mining II Project:**

*Analyzing Data Insights from the IMDb Platform*

### **Authors:**

Bruno Barbieri, Chiara Ferrara, Ankit Kumar Bhagat

*Academic Year 2024/2025*

# Contents

<b>1</b>	<b>Data Understanding and Preparation</b>	<b>2</b>
<b>2</b>	<b>Outliers</b>	<b>2</b>
2.1	COF . . . . .	2
2.2	Isolation Forest . . . . .	2
2.3	ABOD . . . . .	2
<b>3</b>	<b>Imbalanced Learning</b>	<b>2</b>
3.1	Undersampling . . . . .	2
3.2	Oversampling . . . . .	2
3.2.1	SMOTE . . . . .	2
<b>4</b>	<b>Advanced Classification</b>	<b>2</b>
<b>5</b>	<b>Advanced Regression</b>	<b>2</b>

## **Introduction**

### **1 Data Understanding and Preparation**

### **2 Outliers**

#### **2.1 COF**

#### **2.2 Isolation Forest**

#### **2.3 ABOD**

### **3 Imbalanced Learning**

#### **3.1 Undersampling**

#### **3.2 Oversampling**

##### **3.2.1 SMOTE**

### **4 Advanced Classification**

### **5 Advanced Regression**