SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai
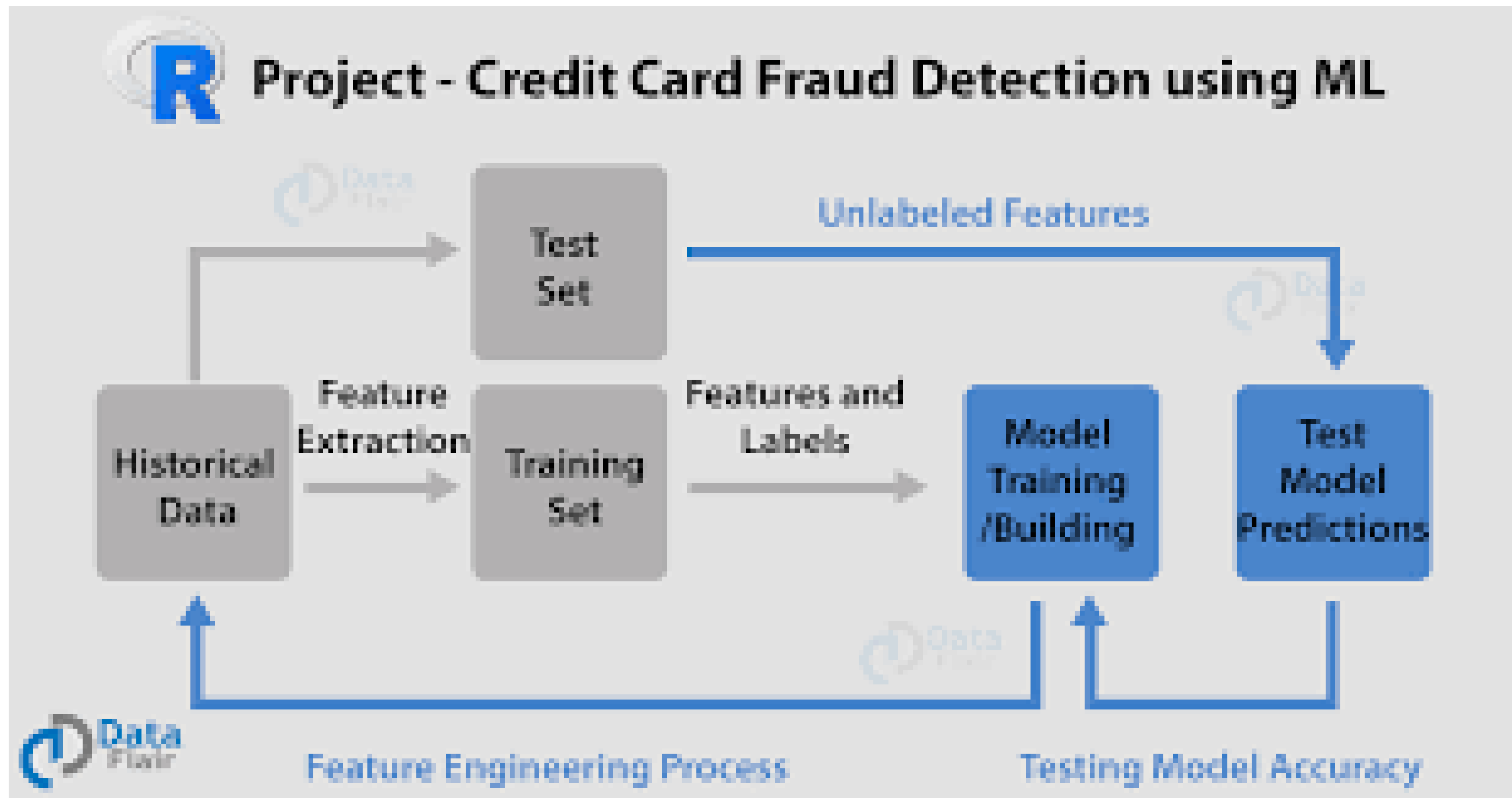
DEPARTMENT OF COMPUTER SCIENCE AND   ENGINEERING

COURSE NAME :19CS407 DATA ANALYTICS WITH R
II YEAR /IV SEMESTER

Unit 1- Introduction

Topic : A Little History on Methodologies for Data Analytics

A Little History on Methodologies for Data Analytics

- ✓ Machine learning, knowledge discovery from data and related areas experienced strong development in the 1990s. Both in academia and industry, the research on these topics was advancing quickly.

- ✓ Naturally, methodologies for projects in these areas, now referred to as data analytics, become a necessity.

- ✓ In the mid-1990s, both in academia and industry, different methodologies were presented

A Little History on Methodologies for Data Analytics

✓ The most successful methodology from academia came from the USA. This was the KDD process of Usama Fayyad, Gregory Piatetsky-Shapiro and Padhraic Smyth . Despite being from academia, the authors had considerable work experience in industry.

✓ The most successful tool from industry, was and still is the CRoss-Industry Standard Process for Data Mining (CRISP-DM) [8]. Conceived in 1996, it later got underway as an European Union project under the ESPRIT funding

A Little History on Methodologies for Data Analytics

✓ In 1999 the first version was presented. An attempt to create a new version began between 2006 and 2008 but no new discoveries are known from these efforts.

✓ CRISP-DM is nowadays used by many different practitioners and by several corporations, in particular IBM. However, despite its popularity, CRISP-DM needs new developments in order to meet the new challenges from the age of big data.

✓ Other methodologies exist. Some of them are domain-specific: they assume the use of a given tool for data analytics. This is not the case for SEMMA, which, despite has been created by SAS, is tool independent. Each letter of its name, SEMMA, refers to one of its five steps: Sample, Explore, Modify, Model and Assess
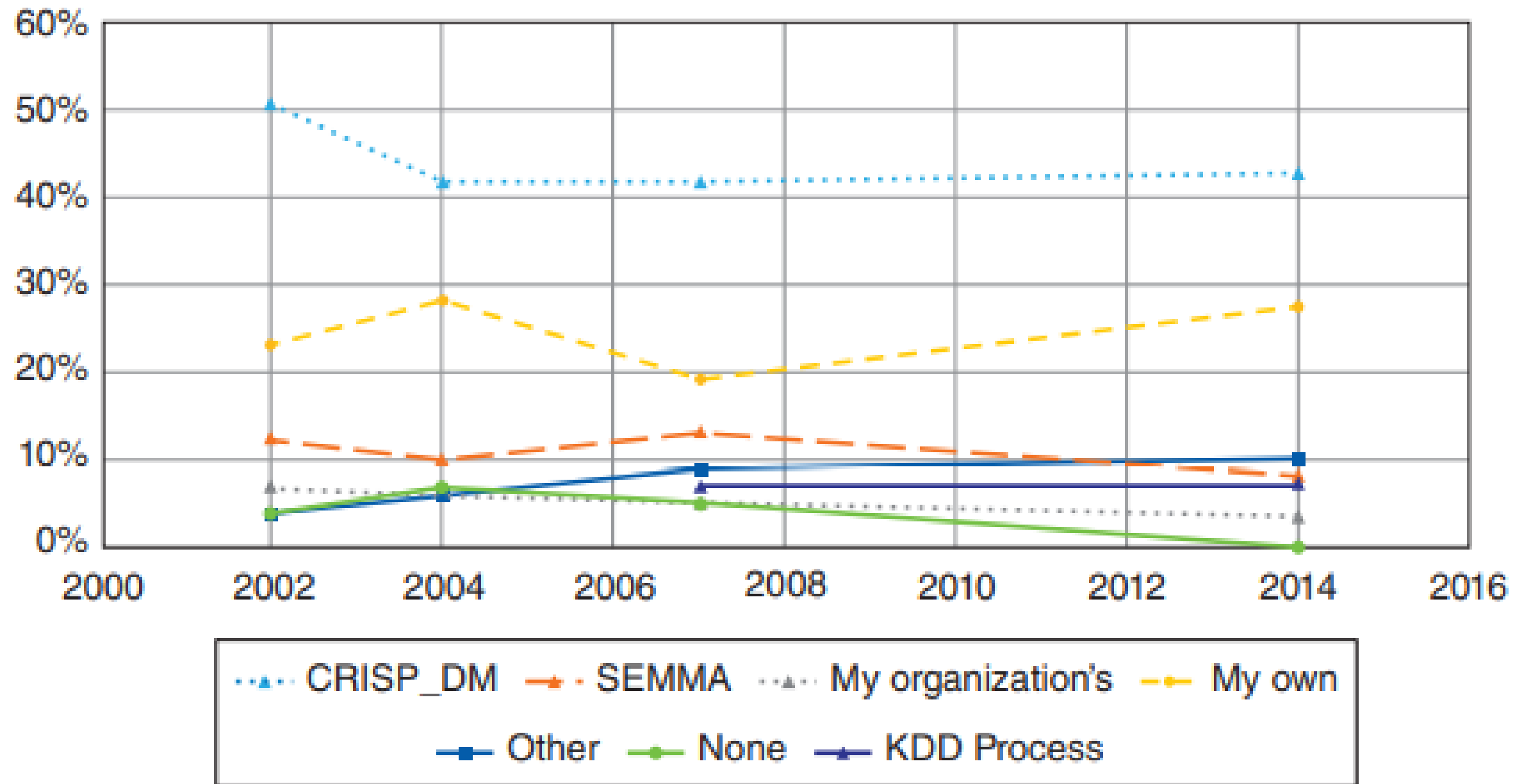
A Little History on Methodologies for Data Analytics

Figure 1.2 The use of different methodologies on data analytics through time.

Assessment 1

To create your own Data Model for above 2 problem

1. João Moreira, Andre Carvalho, Tomás Horvath – "A General Introduction to Data Analytics" – Wiley -2018

Thank You