

CMSC320: Introduction to Data Science

JAMES ZHANG^{*}

January 25, 2024

These are my notes for UMD's CMSC320: Introduction to Data Science, which is an elective. These notes are taken live in class ("live- \TeX -ed). This course is taught by Professor Fardina Fathmul.

Contents

1	Big Data and Data Science Overview	2
----------	---	----------

^{*}Email: jzhang72@terpmail.umd.edu

§1 Big Data and Data Science Overview

The rise of data science over the last 20 years is partially a result of big data.

The 3 V's of big data are

- Volume: the amount of data from myriad sources
- Velocity: the speed at which big data is generated
- Variety: the types of data (structured, semi-structured, unstructured)

In this class, we will also explore the data lifecycle which includes stages such as data collection, data processing, exploratory data analysis, data visualization, analysis, hypothesis testing, machine learning, and insight and policy discussion.