# 10 Data Science - Predicting my preferences

Content Area: **Math**
Course(s):
Time Period: **Full Year**
Length: **5 weeks**
Status: **Published**

## General Overview, Course Description or Course Philosophy

This course combines the study of Statistics and Probability with Data Science. The goal is to have students think critically about data in today's data-driven world and understand its role in the 21st Century economy. Furthermore, students will become familiar with the concepts, topics, and techniques used by data scientists and statisticians in their day-to-day work.

Throughout this course, students will engage in project-based observational studies and experiments to develop their understanding of data analysis, sampling, correlation/causation, bias and uncertainty, probability, modeling with data, as well as making and evaluating data-based arguments. Students will also learn about the roles of data scientists, the power of data in society, machine learning, and how data scientists extract knowledge and insights from real-world data.

In this unit students will be introduced to the big ideas behind machine learning. They will build two different machine learning algorithms to make predictions on whether they will like a song. In this process, they will learn about using vectors and matrices as data structures as well as applying conditional probability and exercising their basic programming abilities. Students will also consider how machine learning impacts their lives and others' lives and will share their newly gained understandings of machine learning with a member of their community. During the unit, students will work in Colab and Edublocks.

## OBJECTIVES, ESSENTIAL QUESTIONS, ENDURING UNDERSTANDINGS

Essential Questions

- How can I build machine learning algorithms that will make predictions on whether I like a song?

Enduring Understandings

- Algorithms can be built based on the preference of music taste.
- Vectors can be used to store data and be represented graphically

## CONTENT AREA STANDARDS

### S.ID

**A. Summarize, represent, and interpret data on a single count or measurement variable**

**B. Summarize, represent, and interpret data on two categorical and quantitative variables**

**C. Interpret linear models**

### S.IC

**A. Understand and evaluate random processes underlying statistical experiments**

**B. Make inferences and justify conclusions from sample surveys, experiments, and observational studies**

### S.CP

**A. Understand independence and conditional probability and use them to interpret data**

**B. Use the rules of probability to compute probabilities of compound events in a uniform probability model**

### S.MD

**A. Calculate expected values and use them to solve problems**

**B. Use probability to evaluate outcomes of decisions**

| | |
|---|---|
| CS.K-12.6.c | Evaluate and refine a computational artifact, multiple times, to enhance its performance, reliability, usability, and accessibility. |
| CS.K-12.7.a | Select, organize, and interpret large data sets from multiple sources to support a claim. |
| CS.K-12.7.b | Describe, justify, and document computational and/or design processes and solutions using appropriate terminology consistent with the intended audience and purpose. |

## RELATED STANDARDS (Technology, 21st Century Life & Careers, ELA Companion Standards are Required)

9.1.12.PB.6: Describe and calculate interest and fees that are applied to various forms of spending, debt and

saving.

| | |
|---|---|
| LA.RI.11-12.7 | Integrate and evaluate multiple sources of information presented in different media or formats (e.g., visually, quantitatively) as well as in words in order to address a question or solve a problem. |
| LA.11-12.SL.11-12.2 | Integrate multiple sources of information presented in diverse formats and media (e.g., visually, quantitatively, orally) in order to make informed decisions and solve problems, evaluating the credibility and accuracy of each source and noting any discrepancies among the data. |
| CRP.K-12.CRP2 | Apply appropriate academic and technical skills. |
| CRP.K-12.CRP4 | Communicate clearly and effectively and with reason. |
| CRP.K-12.CRP7 | Employ valid and reliable research strategies. |
| CRP.K-12.CRP8 | Utilize critical thinking to make sense of problems and persevere in solving them. |
| CRP.K-12.CRP11 | Use technology to enhance productivity. |
| TECH.8.1.12.A.4 | Construct a spreadsheet workbook with multiple worksheets, rename tabs to reflect the data on the worksheet, and use mathematical or logical functions, charts and data from all worksheets to convey the results. |
| TECH.8.1.12.A.5 | Create a report from a relational database consisting of at least two tables and describe the process, and explain the report results. |
| TECH.8.1.12.C.CS2 | Communicate information and ideas to multiple audiences using a variety of media and formats. |
| TECH.8.1.12.C.CS4 | Contribute to project teams to produce original works or solve problems. |
| TECH.8.1.12.E.CS2 | Locate, organize, analyze, evaluate, synthesize, and ethically use information from a variety of sources and media. |
| TECH.8.1.12.E.CS4 | Process data and report results. |
| TECH.8.1.12.F.CS3 | Collect and analyze data to identify solutions and/or make informed decisions. |

## STUDENT LEARNING TARGETS

### Declarative Knowledge

Students will understand that:

- Modeling can be used to predict human behavior
- Algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.
- There are two types of machine learning through a visual: content-based filtering and collaborative filtering.

## Procedural Knowledge

Students will be able to:

- Connect an experience when something was recommended to them to content-based and collaborative filtering and create a visual based on what they think might have happened.
- Choose 3 attributes to work with based on preference and distributions.
- Build algorithms for two different filtering methods in the context of music.
- Build a model with one attribute to predict ratings and test the accuracy of their model.
- Build a model with two attributes to predict ratings and test the accuracy of their model
- Build a model with three attributes to predict ratings and test the accuracy of their model
- Calculate the conditional probabilities that they will like a song, given their partner, likes a song.
- Run the model and consider how well it predicts their preferences.
- Students will use the slides created throughout this unit to share their new understandings of machine learning with someone in their community and reflect on the process of sharing

## EVIDENCE OF LEARNING

## Alternate Assessments

- Portfolios
- Verbal Assessment (instead of written)
- Multiple choice
- Modified Rubrics
- Performance Based Assessments

## Formative Assessments

Observations

Task completion

Student journals and notebooks

Cooperative team work

## Summative Assessments

PBL Assessment

Unit assessments

## RESOURCES (Instructional, Supplemental, Intervention Materials)

You cubed curriculum

Unit 7 resources: https://hsdatascience.youcubed.org/curriculum/

## INTERDISCIPLINARY CONNECTIONS

Educational tech applications

Current Events

Experimentation

## ACCOMMODATIONS & MODIFICATIONS FOR SUBGROUPS

See link to Accommodations & Modifications document in course folder.