PLTW COMPUTER SCIENCE

Activity 3.1.2

Privacy Issues with Data

Introduction

In Unit 1 you learned to program in a variety of languages. In Units 2, 3, and 4, we apply these skills to the technologies that are revolutionizing all sectors of our society and economy:

- Connectivity through the Internet (Unit 2)
- Large-scale data collection and analysis (Unit 3)
- Robotics and physical automation (Unit 4)
- Simulation and modeling (Unit 4)

The changes to our world due to Internet connectivity were evident in the last decade. Data is driving the revolution in the current decade. Data skills (which are a combination of programming skills, math and science content knowledge, a creative eye for graphics, and an ability to explain your results) are the hottest skills on the computing labor market. Information is exploding from everywhere you can imagine, and there simply aren't enough people to work with it.

Data about you can be collected and sold from nearly every purchase, every mouse click, every email or phone call. Do businesses use that data to improve the services they offer you? What do they know about you? As a business owner or employee, how can you use data to do your job better?

Materials

- Computer with access to Internet

Resources

3.1.2 SourceFiles.zip

Procedure

Part I: The New World of Advertising

1. Form pairs as directed by your teacher. Meet or greet each other to practice professional skills. Set team norms.
2. In a previous activity, we examined the issue of privacy and government with respect to law enforcement. Here we examine the issue of privacy and industry. Many retailers track what

you buy and when you buy it. Often retailers buy and sell this information in order to deliver **targeted advertising**, in which you are shown advertisements based on information such as your age, gender, or taste in clothes. For example, instead of paying for an advertisement that will be shown to all viewers, a retailer can buy advertising space within web pages that will be shown to 1000 female students age 15-18 who have a job and who have purchased a sweater in the last six months.

What data might be known by advertisers about you? Brainstorm with your partner.

3. Many web sites sell a **frame** in their web page to advertising companies. A frame is a rectangle of the rendered web page, defined in HTML. The company that buys the frame on the page usually leaves a cookie in your browser so they can identify you later. When you visit a page in which they provide a frame, they can collect data about when you visited that page, when you left the page, and whether you moused over or clicked on anything in their frame. Retailers pay advertising companies for each **impression**, which is the term for a single user being shown a single ad. Advertising companies pride themselves in targeting you accurately. They will charge the retailer more if they get a higher **click-through rate**, which is the percentage of viewers that click on the ads they sell, or **view-through rate**, which is the percentage of viewers that eventually visit the advertised websites even if they don't click on it right away. Advertisers know the age, gender, and at least one retail interest associated with more than 80% of email addresses.

   What companies would want to advertise to you if they knew what you were interested in?

4. In addition to using cookies and IP addresses to track users, retailers and advertising companies collect data about you using **device fingerprinting**, in which the JavaScript in a frame collects device settings and software version numbers.

   The **Electronic Frontier Foundation** is a non-profit organization that advocates for privacy on the Internet. They offer a tool to demonstrate what identifying information your phone, tablet, or computer provides.

   - Visit **https://panopticlick.eff.org/** and select TEST ME.
   - Allowing retailers and other web sites to have information about you does have benefits: they can offer you customized services and information. News, for instance, can be customized to your interests. Why might you want your computer to provide software version numbers to an arbitrary website you visit?

Part II: Privacy Settings and Contracts

5. When you use social media like Facebook, you generate a lot more data than you do when you browse. When you post writing or pictures through a web interface, you might be giving the writing and pictures as property to the company that owns the web service. In addition to these data, the company might collect and sell data about your interactions with other users.

   When you sign up for social media, you are often asked to agree to **Terms of Service** or – in the case of software that you install – a **EULA** (End User License Agreement). These are usually complicated, legally-binding contracts you are signing when you click the Accept button.

   Terms of Service often require you to allow data about you to be collected. The Terms of

Service usually refer to parts of the contract that are contained in additional documents such as a **Privacy Policy** which says whether the company's data about you can be kept forever and whether it can be sold.

Privacy policies sometimes give individual users some control over the terms of the contract by including **opt-in** or **opt-out clauses**. These clauses allow you to customize how much data is collected about you. An opt-in clause will only apply if you take an action like checking a box; the default is that the data collection or service will not occur. An opt-out clause is applied to you by default, but you do have the power to change the contract by taking an action like unchecking a box.

- With your partner, identify the advantage to a retailer of each of these four options:
  - Collect data about customers without giving them individual control.
  - Collect data about most customers but give them opt-out power.
  - Collect data about customers if they opt-in.
  - Do not collect data about customers.
- Summarize the discussion in writing.

6. Examine the Terms of Service, including the contract terms that are included in additional documents like a Privacy Policy, for a service of your choice. It might be a piece of software, your school's email or gradebook portal, an email service independent of your school, or a social media site like Twitter, Facebook, etc. The Terms of Service for a handful of services can be found in `3.1.2.Aa TermsOfService.zip` and `3.1.2.Ab PrivacyPolicy.zip`.
   - What data about you is collected?
   - Is the company allowed to sell data about you? Does this include writing and pictures that you post?
   - Does the contract grant you the right to review or correct the data about you?
   - Does the company promise to keep your data secure?
   - Does the company specify data they will not collect about you?
7. Many social media websites provide an interface for you to modify your **Privacy Settings** even if you have already agreed to share private information in the Terms of Service and the Privacy Policy it includes.

   Privacy settings usually address a very different sort of privacy than the Privacy Policy. The Privacy Policy, which is often part of the Terms of Service, specifies what data the company can collect and whether they can sell it. Privacy Settings, on the other hand, allow you to control what data are shared with other users of the service.

   With your partner, examine the privacy settings for a service like Gmail, Facebook, Twitter, or Pinterest. The privacy settings for a handful of sites can be found in `3.1.2.Ac PrivacySettings.zip`. With your partner, consider how a reasonably private person who wants to use the service ought to set each option. Find at least one option on which you and your partner disagree about the best setting for a reasonably private person. Describe your advice for the person and give your reasoning.

8. The *Consumer Privacy Bill of Rights* was published by the U.S. White House in 2012 but is not enacted in U.S. law and cannot be enforced. The document describes seven rights, a through g, in the excerpts below. For each lettered excerpt, decide whether the Terms of Service, Privacy Policy, and Privacy Settings you examined in the previous steps abide by the “right” asserted by this document.
   - Individual Control

Consumers have a right to exercise control over what personal data companies collect from them and how they use it.

- Transparency

  Consumers have a right to easily understandable and accessible information about privacy and security practices.

- Respect for Context

  Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.

- Security

  Consumers have a right to secure and responsible handling of personal data.

- Access and Accuracy

  Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.

- Focused Collection

  Consumers have a right to reasonable limits on the personal data that companies collect and retain.

- Accountability

  Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.

Part III: Deidentifying Big Data for Beneficial Impact

9. Unlike information that anyone can easily obtain about you, information about your health, education, or finances is considered **sensitive information**. Hospitals, schools, and banks typically encrypt that information and limit who has access to it.

   However, society can benefit if some of the information is public. By using data about millions of people's health, education, and financial behavior, we can improve medicine, learning, and the economy. Scientific knowledge can be discovered lurking in huge data sets by **data mining**, also called **knowledge discovery in data (KDD)**.

   To protect privacy while making data available, information can be deidentified, which means stripping your name, address, and any other content that would connect the information to you. Data can be further deidentified by accumulating groups of people's information together as totals or means. Several laws (like FERPA, the Federal Educational Rights and Privacy Act) treat any data as identified unless it has been accumulated into groups of five or more

people.

Think about your daily experiences with your health, learning, or retail consumption. As a class, brainstorm questions you would be able to answer if you had enough data. Record the results here.

10. Privacy issues still arise even if data are encrypted with limited access. In one case, Target used data mining to discover dddsdfassociations between certain purchases and pregnancy and then applied this knowledge to make a prediction about a customer. Target sent ads for baby products to a teenager because she had bought items typical of a pregnant woman: a larger purse, cotton balls, lotions, etc. Though Target never leaked the specific sensitive information about the girl's purchases, it certainly angered her parents to find out from the retailer that their daughter was pregnant!

    Even if data is deidentified, it can sometimes be **reidentified**. Read the following sections from Chapter 2 in *Blown to Bits* pages 42-48 and answer the questions below. The content is available at **http://www.bitsbook.com/wp-content/uploads/2008/12/chapter2.pdf**

    - *Public Documents Become VERY Public*
    - *Idle Curiosity*
        - Describe how multiple data sets can be combined to extract sensitive information about individuals who were not identified in any of the data sets containing sensitive information.
        - One strategy for protecting privacy is to delete and/or deidentify data over time. How long do you think your school should maintain personally identified data of the following specificity, and what would they use it for?
            - Answers to particular questions from assignments (e.g., your writing for each question here)
            - Summary score for particular assignments (e.g., “87%” in Geo assignment 17)
            - Summary grade for particular courses (e.g., “A” in CSE semester 1)
            - Daily attendance records (e.g., “10 minutes tardy” on April 1, 2015)
            - Summary attendance records (e.g., “3 absences” in 2014-2015)
        - Could your school benefit by keeping some of the data in Part b of this step in a deidentified form? What questions might they be able to answer?

Conclusion

1. Describe some part of the Terms of Service and Privacy Policy from Step 6 that surprised you.
2. The potential risks and benefits of keeping data must be balanced. Write briefly about who you think takes the risks, who reaps the benefits, and who makes the decisions about how long personally identifiable data about you are kept.