Integrating Machine Learning into the UX Design Process

by

Tauhid Ahmed

December 2018

Presented to the

Division of Science, Information Arts, and Technologies

University of Baltimore

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

Approved by: ________________________________

Greg Walsh, Thesis Advisor/ Program Director

____

Bridget Blodgett, Committee Member

Abstract

The following paper discusses how machine learning is becoming the new user experience tool for designers. Throughout the last decade, machine learning has vastly improved the user experience and human-machine interfacing by permitting machines to learn who we are and how we would like our systems to communicate with us. Machine learning opens the door for user interface and user experience design opportunities that could further meet users' needs. To explore this phenomenon, *Coupon Buddy* was designed using a prototyping strategy to explore how machine learning could classify comments and adapt to user interaction and feedback. More specifically, the application functioned as a research channel to observe how UX designers could improve design processes for better user experiences through the accumulation of machine learning. *Coupon Buddy* was designed to allow users to save all their coupons in one place and use it for their shopping needs. Not only did the creation of *Coupon Buddy* prototypes allow us to investigate how much knowledge of machine learning our participants already had, but it facilitated ideas for how machine learning corresponds to a stronger UX design approach.

Acknowledgments

I want to thank my friend Marc, who has helped since the beginning for this entire thesis project. Without his help and motivation, I wouldn't have reached this milestone. I would also like to thank my close friend Deswyn, who helped me brainstorm ideas about what kind of prospective machine learning application to propose that could expand and accommodate a discussion for machine learning abilities. I want to thank my parents and sister for their support; they believed in me to complete this entire project and program. Without their love and support through all these years, I would have never come to an end. Let's not forget about my great advisor Greg Walsh, who has supported and mentored me the whole way on how to balance this research project. Without his help to structure my research methods, the project probably would have never happened.

Table of Contents

List of Figures

Chapter 1: Literature Review

Introduction

User experience (UX) has transformed how we build applications today. Human Computer Interaction (HCI) researchers have dived into methods to understand how the rise of machine learning is the future for UX experts, "HCI has become particularly interested in using machine learning (ML) to improve user experience (UX)." (Yang, Banovic, & Zimmerman, 2018) Machine learning is advancing rapidly due to the sophisticated systems we are developing today. It has reformed user experiences for smartphones, online retailers, games, and interactive systems. Machine learning makes accurate decision algorithms by assimilating user behavior and interaction. Machine learning has granted computers the capability to do the extraordinary and to mimic our human brains by observing the data that we interact with. With the growing popularity of machine learning and AI (Artificial Intelligence), how can we help UX designers integrate the practice of machine learning into their design thinking to reduce redundancy and the conventional static design process, thus encouraging them to become better designers? The *Coupon Buddy* prototype strove to explore this question by examining how machine learning classification of user comments and user feedback could be applied to application development within UX design.

How Machine Learning Practice is Evolving Within UX

UX data research experts are delving into the debate on how UX designers can integrate machine learning (ML) into design thinking. "UX designers struggle to understand the capabilities and limitations of ML… They typically joined projects towards the end, after the functional decisions had been made." (Yang, Scuito, Zimmerman, Forlizzi, & Steinfeld, 2018)

While this may be true, UX designers tend to rely on usability test workshops through recruited participants, which requires time, money, and effort to collect recommendations for future improvements. “Recently, design researchers and educators began taking actions to address this problem. A few have developed designer-focused education materials, meant to teach the technical concepts of ML.” (Yang, Scuito, et al., 2018)

Ideally, this will not just guide UX designers to enhance product design with positive experience, but the collection of user responses can predict future outcomes. Thus, designers will have the proficiency to understand the flaws of their product design, inheriting the guidance from machine learning.

Challenges and Opportunities for Machine Learning

Dove et al. have researched what could be done to help UX designers and machine learning experts collaborate. Dove explained “ We wanted to understand whether the difficulty of working with ML as a design material might be inhibiting the design response to this not so new technology.” (Dove, Halskov, Forlizzi, & Zimmerman, 2017) Their approach was to conduct a survey to better understand each participant if they ever had been involved in ML design practice or research. (Dove et al., 2017)

Furthermore, to gain additional insights about projects, they were specifically asked if they ever teamed up with experienced ML data scientists. If projects involved machine learning, significant challenges during the projects had to be identified. Although ML mechanisms were mentioned during the project, “…respondents listed features such as personalization, customization, prediction, and recommendation…”(Dove et al., 2017), in general, it remained true that integrating machine learning for applications remained complex and challenging.

In addition to Dove et al’s research on what could be done regarding the lack of training for adapting machine learning concepts to design methods, other researchers have shared frustrations where machine learning concepts are not taught to current user interface (UI) designers to learn from and adapt their algorithms. As Yang et al.

mentioned, “Practice-focused design patterns as well as books and articles on wireframing do not discuss how to design for adaptation or how to collect the information needed for ML.” (Yang, Zimmerman, Steinfeld, & Tomasic, 2016.)

Despite the adaptation of machine learning in UX, Stumpf explained “Users were willing to provide a generous amount of rich feedback to machine learning systems, and that the types of some of this rich feedback seem promising for assimilation by machine learning algorithms.” (Stumpf et al., 2008) As a result, the development of a machine learning tool would not only evaluate the patterns of end users’ responses, but at the same time, feedback should be utilized to teach machines where to make alterations during the design process. “It has become feasible to allow these systems to continue to adapt to end users by learning from their behavior after deployment.” (Stumpf et al., 2008) Additionally, by increasing use of machine learning in the design method, Yang believes “data scientists are the next UX designers.” (Yang, Scuito, et al., 2018) Data scientists can help UX students obtain data when designing ML products. As Yang explained “Student designers entering industry now and in the near future will need to work with data scientists in this unstructured manner, with very few boundary objects to help scaffold their collaborations.” (Yang et al., 2018)

A group of researchers have made an emotional prosthetic machine learning intelligence system, which allows users to share their emotions over time. Emotion is relevant to understanding users’ needs or how they want the system to behave. McDuff et al. designed a machine learning system *AffectAura* and explains what the system does: “We designed a multimodal sensor set-up for continuous logging of audio, visual, physiological, and contextual data, a classification scheme for predicting user affective state, and an interface for user reflection. The system continuously predicts a user’s valence, arousal and engagement, and correlates this with information on events, communications and data interaction.” (McDuff, Karlson, Kapoor, Roseway, & Czerwinski, 2012) In addition to how system observes users’ emotions by assessing the dimensions of valence, arousal, and engagement, it is essential to acknowledge how a user's emotion contributes a relevant feedback response during the design process. The ability to learn users' emotions during interaction with a system is critical to training effectively-responding machine learning systems. Consequently, deriving emotions from users' reactions and teaching systems how users feel about design is what makes a successful product or service.

Teaching Language Platforms Influenced by Machine Learning

Research has shown that essential language skills critically influence the design of applications. Therefore, when collecting user responses, it is important to observe that a collection of words will have a critical impact on training machines in the patterns of various words used to provide context and both subjective and objective meaning. This will allow UX designers to distinguish language among common terms and related terms.

"According to the International Dyslexia Association (IDA), as many as 15 to 20 percent of the world population is detected with the symptoms of dyslexia and affecting people regardless of gender, different ethnic and socio-economic backgrounds." (Hamid, Admodisastro, & Ghani, 2015) The Malaysian Ministry of Education reported that "28,139 students have registered for special education who need help to read, write, and have learning difficulties." (Hamid et al., 2015) Under those circumstances, the school has cooperated with dyslexia experts in order to create a computer-generated machine learning model that would benefit students. Consequently, the designers proposed three efficient design models with different functions and needs in order to facilitate the learning process. The designed systems are *MyLexics*, *Bijak Membaca*, and *Dyslexia Baca. MyLexics* "helps children in reading and writing using dual coding theory and scaffolding teaching strategies, which enhanced word or letter recognition through visual and verbal format." (Hamid et al., 2015) "*Bijak Membaca* has combined several phonic reading techniques include letters, syllable, word and sentence using a multisensory approach." (Hamid et al., 2015) Lastly, *Dyslexia Baca* "used multisensory approach in teaching special orthographic such as distinctions of 'b','d', 'p','q', and 'm','n'." (Hamid et al., 2015) The motivation to build these systems is a lesson in best practices in incorporating human language skills into machine learning because "a learning model that uses the ML approach has an ability to recognize student's learning strategies, detect student affect while learning, predict student performance, and act as a tutor to make teaching decisions." (Hamid et al., 2015)

Data Labeling Machine Learning

Although language has fundamental significance when assembling data to teach machines the language utilized by end users or search terms, it is relevant to educate machines where and how to label data. Teaching systems how to label and sort data is one aspect of this fundamental process. For instance, when training machines to assess the list of words used during crowdsourced feedback collection, research clarifies "One task necessary to develop any machine learning system is the provision of labeled data, i.e. data that has been pre-classified typically by a person who examines and labels data to generate a sufficient dataset for the classifier to learn from." (Sun, Lank, & Terry, 2017) In certain situations, especially during the system design process, participants would recommend which action cues ("next, back, forward, refresh, save, etc.") would function best based on their knowledge in making the system intuitive and navigable. For instance, navigation menu labels, button labels, and wayfinding are essential features to consider during a user research session. Not only will this routine advise designers to track the ML process during application creation, but training machine learning on the sense of alternative terms can further predict and recommend which words have deeper meaning for accurate results.

More specifically, researchers are conscious about the challenges UX designers or developers can encounter when facing the circumstances for data labeling. Research was conducted by Sun et al. to propose a labeling system named *Label-and-Learn*. Sun et al.

illustrates how the machine learning system supports designers and developers by declaring "*Label-and-Learn* is a system designed for developers who want to train a machine learning classifier to solve a specific problem. It can help them estimate the classifier's expected performance at the early construction phase of machine learning during labeling." (Sun et al., 2017)

Recommendation Strategy in Machine Learning

Another emerging functionality for machine learning is the recommender system. The recommendation strategy has evolved tremendously by high-level improvements led by Amazon and Netflix. Design experts are researching ways to utilize this method in other fields within the machine learning design strategy. At the same time, user interface design experts are constantly exploring how recommendation strategy would help them cultivate training systems to obtain quality feedback responses and learn users' collaboration behavior. According to a research studied by Lee et al., "Collaborative filtering algorithms focus on similarity among users or similarity among items using user's ratings… However, many users assign arbitrary ratings that do not reflect their true opinions." (Lee & Park, 2008.) To circumvent the process of collecting feedback, their research focused on developing a model that would alternate how data will be collected by filtering out arbitrary submissions. More specifically, they sought to conduct a research study which contained a collection of review responses, recommendations, and ratings that were considered explicit feedback. (Lee et al. 2008) The research results proved that if data are to be collected by adapting the method for implicit feedback, the results from recommendation collection had a higher frequency of accurate recommendations. Applied lessons from these studies have significantly improved the way e-commerce websites study end user shopping patterns, browsing behavior, or search results to find related products.

Now, design experts are exploring opportunities how machine learning recommendations could lead to a design improvement strategy. Research by Chan, Stone, and Szeto et al. has shown that the steep learning curve for data processing and learning algorithms is one of the biggest challenges for software developers to use machine learning in building predictive applications. To tackle this challenge, Chan et al. have designed a system *PredictionIO*, which is "an open source machine learning server that comes with a step-by-step graphical user interface for developers to evaluate, compare, and deploy scalable learning algorithms, tune hyperparameters of algorithms manually or automatically, and evaluate model training status." (Chan et al., 2013)

This approach has not just benefited the research team personally, but actually, by delivering a proposed solution, it has prominently facilitated developers elsewhere not to spend too much effort redundantly analyzing their own data. Furthermore, Chan et al. concludes "Our system eases the challenges of building predictive software applications with machine learning." (Chan et al., 2013)

Crowdsourcing Design Feedback in Machine Learning

Crowdsourcing is used widely by software developing employees or even product managers to distribute their work to the public and gather feedback for design revisions. A more formal definition is "…a method of distributing work to a large number of workers (the crowd) both inside and outside of an organization, for the purpose of improving decision making, completing cumbersome tasks, or even co-creation of designs and other projects." (Chiu, Liang, & Turban, 2014) According to research done by Krause et al, collecting design response feedback can be challenging when crowdsourcing design feedback *details.* "A challenge of crowdsourcing design feedback is that the results are often low quality. The reason for this may include diverse contributors who lack sufficient motivation, context, knowledge, and sensitivity to provide effective feedback."

To circumvent this predicament, the team has proposed a natural language model that automatically extracts language features that corresponds with ratings of perceived helpfulness. (Krause et al., 2017) First, the research team's approach was to find out how they could help users write effective feedback which would also be understood as helpful feedback. The key was to provide directions and instructions by illustrating a *critique style guide* which can help participants generate feedback to acknowledge the template designers are presuming to obtain from the response. “To identify the features perceived as most helpful, we conducted a linguistic analysis of the writing style of the collected feedback.” (Krause et al., 2017) Language was not the only relevant way to analyze and observe how participants would provide feedback on designs. Moreover, the method sought to justify if designers provided sufficient directions in controlled guidance on how to write structured comments and whether the insights from users’ comments had greater benefits to improve the design.

In addition to this research conducted by Krause et al., a similar scenario would occur when collecting users’ responses to observe how end users utilize natural language in detail. Therefore, it is relevant to teach machine learning how natural language can assist UX designers when suggesting recommendations for their application design.

The Importance of Data Categorization in Machine Learning

Data categorization is relevant to teach systems how they should classify user feedback responses. If machine learning systems can inherit the algorithms to categorize user feedback responses in a consistent pattern, the accuracy of the system can help UX designers detect that declared terms have been used in previous user feedback sessions. As a result, design experts can acknowledge patterns in how many times the same suggestion has been mentioned before and prioritize tasks for improvements.

Tsukada et al.’s team has conducted research and designed a smart mail delivery system *LetterTwitter,* “a smart mailbox that can detect the arrival of s-mail, distinguish letters from flyers, and notify users of their arrival.” (Tsukada, Mizushima, Ogata, & Siio, 2010) This research shows how every letter is separated by their data classification. Not only did it demonstrate how the significance of data categorization is relevant for machine learning systems, but it showed how keeping data organized makes tasks simpler.

In addition to the research conducted by Tsukada and his team, it is critical for UX designers to acknowledge how advantageous it can be if systems could detect words that have similar meanings. Therefore, strengthening machine learning pattern recognition and categorization techniques will lead to user-responsive UX design improvements.

The articles discussed here demonstrate how machine learning is the present and future approach when designing applications, services, or products. As mentioned in the introduction, data scientists are encouraging UX students, candidates, or even specialized design experts who have been in design practice for years, to imbibe the practice of machine learning. UX designers teaming up with data scientists would not only gain knowledge in machine learning, but their ability to obtain quality results would improve their iterative and responsive design process.

By connecting these examples and lessons learned, our method will target how we can develop a test application that responds to and learns from diverse feedback, recommendations, or other suggestions that can help improve our application design, particularly in the realm of comment/feedback classification. We'll consider how machine learning can create user-centric products by honing responsiveness to users' behavior. Our goal is to explore strategic analysis to consider how well our design will perform in terms of intuition, language and ease of use, label coordination, and most importantly, if our design "looks and feels" complete. As such, by gathering all the data and feedback necessary from our participants, we will study and develop recommendations for how we can integrate machine learning functionality into our design process to efficiently classify user feedback/comments and in turn respond to each user's pattern of behavior.

Chapter 2: Research Methods

Overview

As discussed in the literature review, machine learning plays an important role for enhancing the user experience by building smarter applications, services, or products. This section will therefore examine further how we structured our method of investigation and research procedures to apply machine learning's best practices to classify comments and improve our design. As researchers, our goal was to collect as much responsive data as possible through the feedback, recommendations, and suggestions obtained from our recruited participants. In the next chapter we'll discuss how this research helped us learn from our participants to deliver a more insightful user experience by proposing learning patterns and classification approaches with machine learning potential to apply to a revised prototype.

Participants

Our goal was to recruit 100 participants through social media platforms, referrals, and by word of mouth. We ultimately engaged 83 participants, between whom we obtained 239 unique pieces of feedback. Participants had to be 18 years old in order to complete the survey. Participants were also asked to describe their current field of work after providing feedback for the prototype application *Coupon Buddy*. More specifically, our survey approach provided controlled guidance by asking for each mockup:

- **What needs to be improved?**
- **What would you add to it?**
- **What do you like about it?**

At the end of the screen-by-screen feedback, each participant was shown all of the screens and asked short questions about how far along the design was and how easily changed it could be.

Limitations for Survey Approach

Although gathering the feedback for the designs was carried out through a survey, we were well aware that survey responses often have limitations in their lack of reliable or thorough responses. With this in mind, our questions were formatted in the manner of both open and closed questions. The former format was used in instances in which more complex responses were desired, while in the latter format respondents had a fixed number of responses to choose from. For instance, the structure of questions was limited to yes/no choices once the participants had completed responding to questions referred to design improvements. The decision behind assigning open-end questions for collecting feedback resulted in a stronger approach for both quantitative and qualitative analysis. For example, it helped us gather better qualitative data by observing the prototypes. In particular, our main goal and objective was carried out to evaluate what participants thought about the prototypes. Ideally, open-ended questions helped participants provide thorough explanatory insights, in which they had the opportunity to express as many suggestions and details as possible.

A Quick Overview of How Coupon Buddy Works

Coupon Buddy is a fictional project created to elicit feedback to train machine learning applications about feedback. It is designed to be an application that enables coupon lovers to store all their coupons in one place. After a user installs *Coupon Buddy*, the first thing they do is create an account: a user can sign up using their Facebook account or their email address. Once an account has been created, the first menu page displays the list of functions the application can perform. To add a coupon, a user clicks on "Add Coupon." At this point the user fills in all the necessary information about their coupon before saving it digitally.

Once a coupon has been added, the user has the option to review the coupon by going to "My Coupons." Not only can the user review their newly-added coupons, but the application also categorizes each coupon by color code: If the coupon displays green, it is still valid. If it's yellow, it's close to expiring. If it's red, the coupon has already expired.

Finally, if the user would like to adjust application settings, "Preferences" allows the user to control how to arrange coupons. Here a user can manage coupons by filtering collections and automatically adding similar coupons based on what other coupons were added earlier. *Coupon Buddy* has the potential to track a user's activity and suggest similar offers and savings opportunities.

Procedure

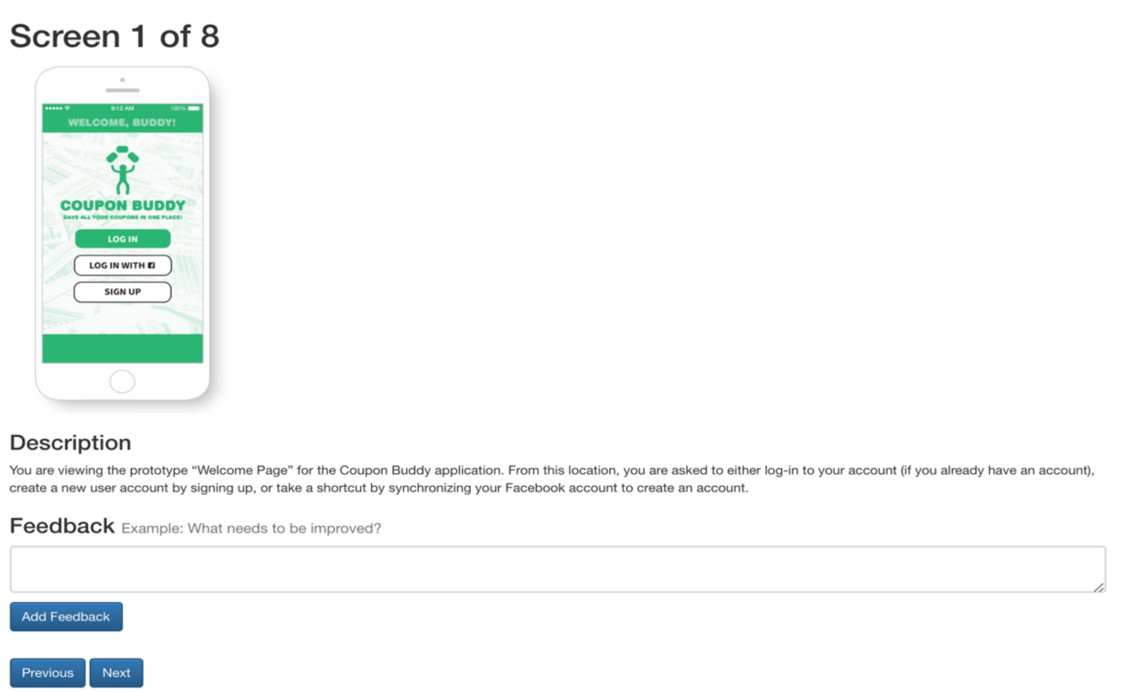

Figure 1: Feedback Analysis Format

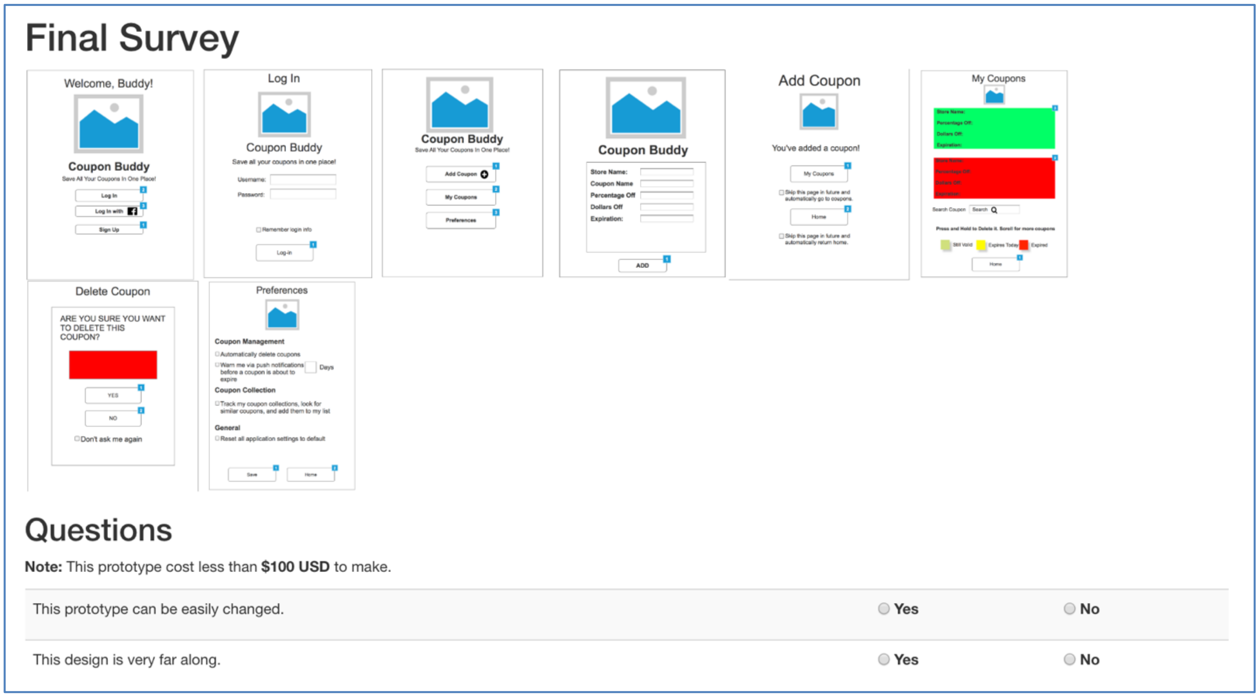

Figure 2: Final Questions

To help us measure and gather the data for the prototype model *Coupon Buddy*, data was collected using a web-based survey database system illustrating sets of prototypes. Each user was assigned a condition of paper prototypes, low-fidelity prototypes, or high-fidelity prototypes. The assigned tasks were to read the scenario description under each prototype and provide feedback related to the random generated question asked about the prototype (Figure 1). After completion, participants were asked succinctly if the prototype could be changed and to specify if the prototype was far along in the design process (See Figure 2). Before participants were assigned to review and comment on the design models, an informed consent form was supplied to each user to obtain their consent that their proffered feedback would be kept confidential and that the collected data would ideally help us, as the user experience researchers, teach machines the algorithms necessary to create predictions for designed interventions.

There were three different versions of prototypes for the *Coupon Buddy* application for participants' review (see Figures 6 through 8). The first version of the prototypes were paper prototypes which were clearly hand-drawn, the second version were digitally created using the software Axure, and the third version were designed using Adobe Illustrator, the last of which depicted in the most stylized and polished manner the interactive graphical elements that participants would be able to interpret as a "final" or "realistic" application.

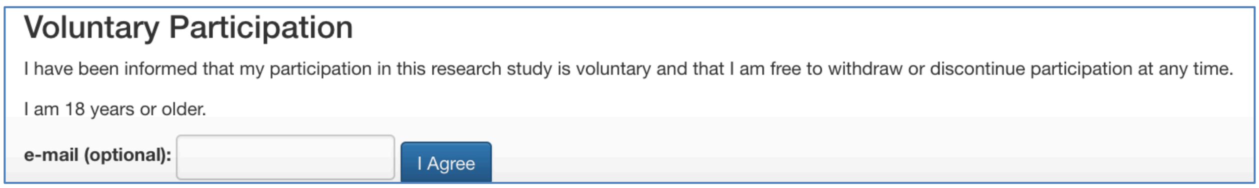

Figure 3: Email Address for Contacting Participants

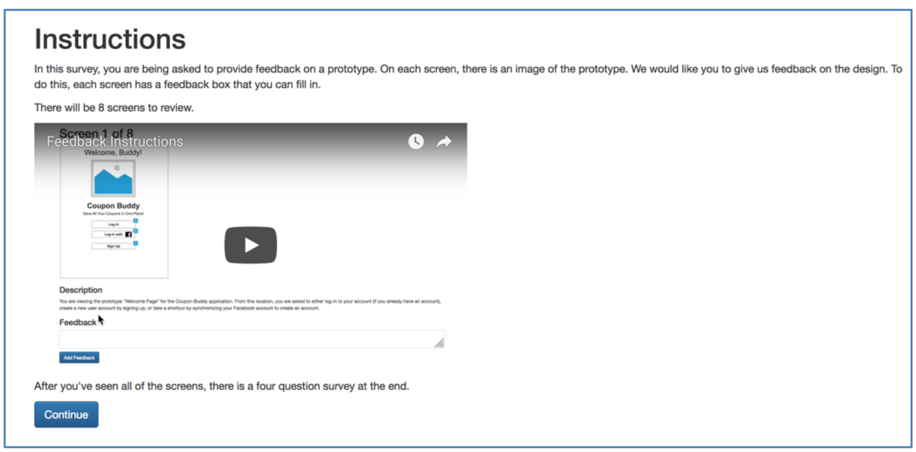

Figure 4: Instructions on How to Take the Survey

The participants' first task was to optionally enter their email addresses (Figure 3) so we could acknowledge if the participants would need to be contacted again, or if any comments supplied by participants required follow-up or clarification. After emails were collected and entered into the database, the follow-up displayed an instruction tutorial

(Figure 4) describing the purpose of the research and what kind of feedback we were trying to obtain from the participants. Prior to displaying the prototypes, we on the back end assigned a condition statement to select which versions of the prototypes (Paper, Axure, or Illustrator) would be consistently and uniformly displayed for each participant throughout the survey. For example, if the initial selection condition was set for a paper prototype, then all subsequent survey imagery would only call from and display additional paper prototype images for the participants to provide feedback and recommendations.

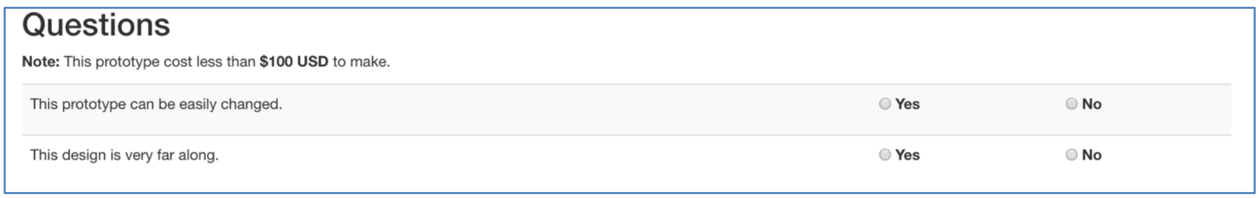

Figure 5 Format for Questions

The format of the questions in the survey was kept simple so participants could answer without hesitation or confusion (Figure 5). As researchers, we were not present before the participants for face-to-face interaction and clarification, so questions weren't phrased in a difficult or complex manner, nor did they prompt participants to supply excessively-complex feedback. We were aware that since this study was conducted through an online portal with minimal assistance, it was necessary to keep the format of the questions as clear as possible.

Prototypes

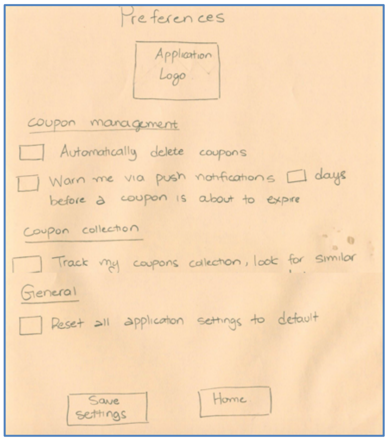

Figure 6: Paper Prototype Preferences

Figure 6 demonstrates the paper prototype model for the preferences page of *Coupon Buddy*. Participants were asked to review the features and functions within this design iteration and evaluate if it seemed intuitive. Their task was to complete the survey questions by providing as many insightful comments as possible.

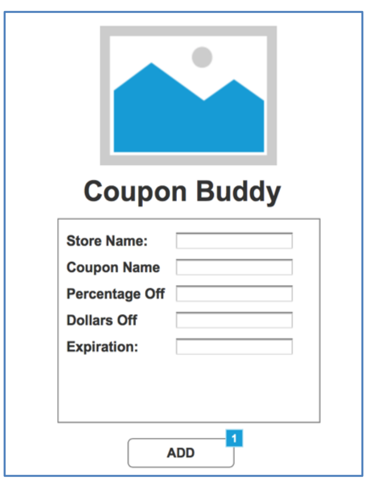

Figure 7: Axure Prototype Add Coupon

Figure 7 demonstrates the Axure prototype for the process of adding a coupon. The participants' tasks were to understand if this page looked intuitive, complete, or to prompt them to suggest additional features.

Figure 8: Adobe Illustrator My Coupons

Figure 8 illustrates the most polished prototype for adding a coupon into *Coupon Buddy*. At this stage, the participants were looking into the feature where coupons are saved into their account once added. The colors for each coupon were color coded by their expiration dates. If a coupon was still valid, the coupon would stay green, but if expired it would turn red. It was necessary for each participant to provide as much feedback on the completeness of the functionality and appearance with regard to potential opportunities for integrating machine learning ideas.

As we have discussed, participants were recruited to help our application *Coupon Buddy* adapt to user behavior through the classification of comments and similar pattern recognition, and we strove to make the survey approach efficient enough to collect as much relevant feedback as possible. We chose to conduct the study through a survey precisely to help us explore our participants' thoughts and recommendations on integrating machine learning ideas, and to examine any patterns between responses to see if those patterns could be exploited for machine learning classification techniques. Given these points, we believe that the survey method was ideal to hone our participants' attention to improve our design. In the next chapter, we will further analyze the feedback results we collected into our database. The feedback obtained from each participant's response helped us analyze reactions to each prototype, but it also helped us understand how people would want *Coupon Buddy* to work if machine learning practices were integrated into the application.

Chapter 3: Research Feedback Analysis

Overview

This chapter will review the feedback and suggestions that our participants gave for the *Coupon Buddy* prototypes. The feedback results, which were collected through our online database survey system, were reviewed to analyze how our participants responded to different versions of the prototypes, and to examine patterns between responses for further exploration of machine learning classification. Participants provided many insights for adding functionalities that could integrate machine learning methods into *Coupon Buddy*.

Participants Survey Collection Database

+ Options

←T→	id	user_id	feedback	page	type	timestamp
Edit Copy Delete	259	141	0:Clear+instructions%2C+action+button+is+large+and...	addpage	paper	2018-10-18 12:49:30
Edit Copy Delete	260	141	3:option+to+scan+the+coupon+or+take+a+picture+so+t...	addpage	paper	2018-10-18 12:50:03
Edit Copy Delete	271	142	3:This+is+a+lot+of+information+to+input+manually.+...	addpage	paper	2018-10-18 14:22:12
Edit Copy Delete	272	142	3:The+logo+could+be+smaller.+I+don%27t+think+logos...	addpage	paper	2018-10-18 14:22:52
Edit Copy Delete	279	143	1:It+feels+a+little+tedious+to+have+both+percentag...	addpage	tech	2018-10-18 16:51:33
Edit Copy Delete	287	146	0:Same+as+earlier%2C+fields+are+a+tad+bit+hard+to+...	addpage	graphic	2018-10-23 13:28:32
Edit Copy Delete	295	147	0:Contains+all+the+relevant+info+for+a+coupon.+The...	addpage	graphic	2018-10-23 23:41:28
Edit Copy Delete	261	141	2:%22Add+another+coupon%3F%22+option+	couponadded	paper	2018-10-18 12:51:16
Edit Copy Delete	262	141	0:option+to+skip+this+page+	couponadded	paper	2018-10-18 12:51:25
Edit Copy Delete	273	142	1:1.+Home+and+My+Coupons+should+be+one+in+the+same...	couponadded	paper	2018-10-18 14:28:16
Edit Copy Delete	280	143	1:I+wonder+if+you%27d+need+this+page+if+you+added+...	couponadded	tech	2018-10-18 16:54:21
Edit Copy Delete	288	146	0:Options+are+nice+	couponadded	graphic	2018-10-23 13:29:03

Figure 9: Database Server for Survey Responses

This figure illustrates our web-based database server that was used to collect our survey responses from participants. Each feedback entry articulates different suggestions and comments for the different versions of the prototypes. The figure provides an example of entries filtered for the “add page” and “coupon added” prototypes. The data shows the feedback about what kinds of suggestions were recommended to make the application functional.

Sorting the data by prototype pages was essential for data analysis, because it significantly helped us observe how many responses were given for a specific prototype. Having the ability to count the number of responses for each page allowed us to observe similar suggestions and calculate the frequency of similar feedback.

Prototype Survey Results Analysis

As our participants were randomly assigned to a specific version for our *Coupon Buddy* prototype, we were surprised to learn that our participants' responses were not influenced by the three different versions of prototypes – paper, Axure, or Illustrator - that were assigned for them to review. That is, the recommendations and suggestions they provided were clearly content-oriented rather than appearance-oriented. We have categorized our analysis, examined patterns for machine learning exploitation, and created bar charts to illustrate what additional improvements were suggested. Each different figure displays a common or repeating point of feedback per prototype page.

Figure 10: Analysis for Login Page

Figure 10 reflects our participants' comments about the login page. Each number represents how frequently participants mentioned the same suggestions during the survey. Surprisingly, and as depicted in the first column, a majority of the responses (6) heavily debated the logo for the application. Participants thought it was unnecessary or irrelevant to display the application logo when a user would need to sign into the application. As one of the participants mentioned "Making the logo smaller to allow the user inputs screen appear above the onscreen keyboard would be helpful." This response helped us acknowledge that users wanted more significance given to the text fields where users could insert their login information. The visibility for the logo brand needed less attention for the application's purpose.

Three participants felt it would be nice to add features such as Touch ID capability for login. On the other hand, participants were also concerned if they could use their email address as their username instead of remembering their username all the time.

Furthermore, participants mentioned that it would be ideal if the text fields for inserting user credentials could be changed to white instead of being transparent green. Participants felt that it could easily complicate users to differentiate between the text labels and input fields.

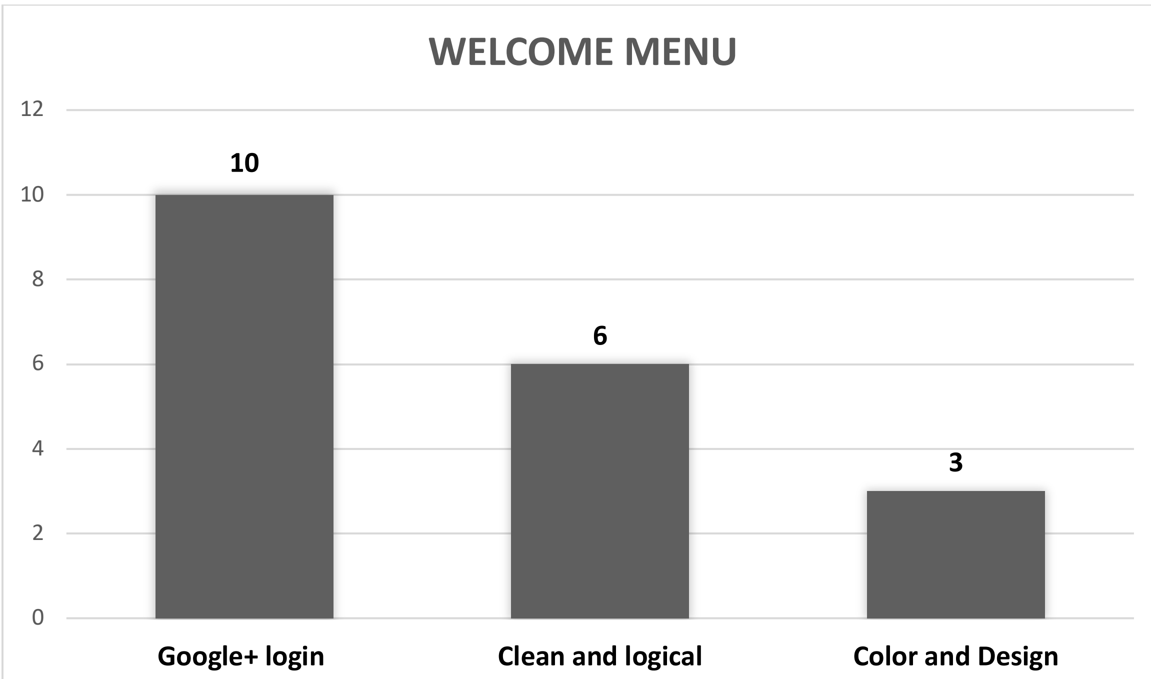

Figure 11: Analysis for Welcome Page

Figure 11 demonstrates what kind of suggestions were most common for the welcome page. Participants didn't have trouble understanding the details, design, and layout; the majority of the responses (10) were linked to adding features in which users could log in with Google+ accounts. As one of the participants mentioned "Perhaps login in with Google+ would also be relevant for users who are reluctant to link their social media with an app for fear of posting without user consent." Three participants recommended design layout plans about how they would like to see the buttons grouped in order. For instance, some participants mentioned, “Group the two login buttons together and add more white spaces between login signups.” Conversely, others recommended the design move the buttons away from the logo to provide more emphasis to the logo.

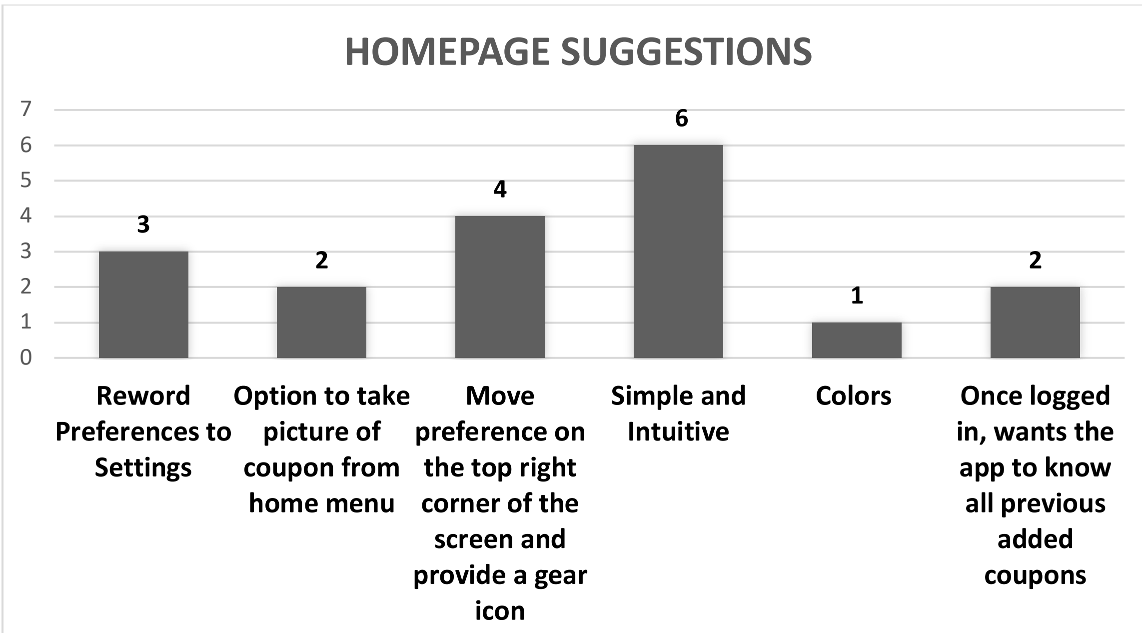

Figure 12: Data Analysis Homepage

Figure 12 reflects a wide variety of feedback for the *Coupon Buddy* homepage, which is where user interaction and the resulting feedback unsurprisingly becomes more complex. Although the homepage seemed intuitive upon login, there were mixed results in how users wanted the layout to behave once logged in. Two participants recommended they would appreciate a function to have their previously-added coupons displayed in the home menu, and other two participants requested adding an extra button for taking coupon pictures. This is a potential field ripe for machine learning application: no one layout or order will satisfy everyone, so the order/layout of options could adjust to user behavior by sorting in the order of which buttons/functions are used most frequently depending on each unique user's behavior.

Additionally, four participants preferred moving the "preferences" button to the top right corner of the screen, reflecting the importance and influence that design standardization has had on app users over the last decade. Many participants believed we are conditioned to look either to the top right or left corners with gear or "hamburger" icons to adjust settings: "Replace the preferences button with a cog icon and position it to the top right in the navigation bar." Three participants were concerned about the language used for the term "preferences." There were suggestions to rephrase "preferences" with "settings," which is more of a universal term used to search for account adjustments. In summary, after analyzing how participants provided a wide range of insights to add additional options to brighten the menu, nearly half of participants still believed that the options available were still intuitive and easy to use.

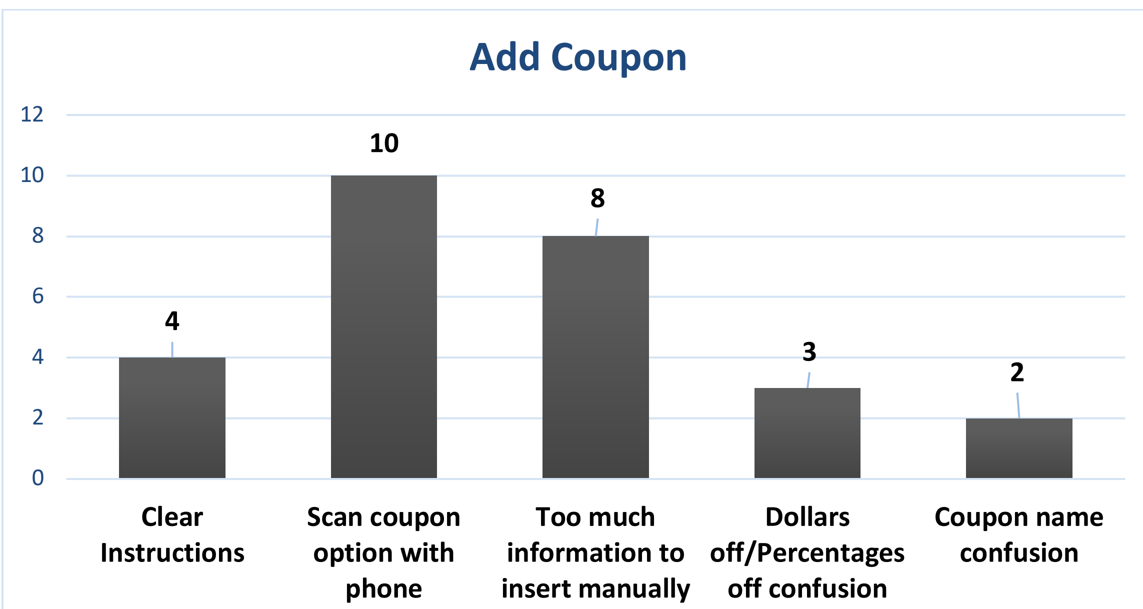

Figure 13: Add Coupon Prototype

This prototype for adding a coupon into the system sparked a rich debate over how users would have to enter their coupon information to store and save the data for later use. A majority– 23 participants– were unsatisfied with why they would have to manually type all the coupon details themselves instead of having a camera feature where the application could scan the coupon and auto-populate the necessary fields itself (this is perhaps the most fruitful field to examine for machine learning in future iterations of the app). The most common feedback was “Will you be able to scan a coupon with your phone and automatically populate the information?”

Contrarily, some of our participants responded positively to the layout by complimenting the layout for manually entering the coupon details, responding “This should be a manual entry page in case the scanner fails to import the coupon info.” Lastly, a number of participants were confused by the meaning of the phrase “coupon name.” One of the participants asked “Is ‘coupon name’ the brand of the coupon? Make this clearer.” In the future, during machine learning language labeling training sessions, we will examine how to distinguish and decide if the system will be trained on coupon names or brand names when entering coupon details.

At the same time, three participants were confused how inserting information about dollars or percentages functioned in the app. Some of the participants mused “I presume if you entered percentage off, then dollars off will be greyed out inaccessible and vice versa.” Nevertheless, we intended the app to allow a user to enter either or both at the same time, but the questions raised were legitimate. A tip from one of the participants suggested “Perhaps add subcategories for percentages off and dollars off for minimum purchases.” We believed that implementing these options during machine learning training could definitely reduce inconsistencies within *Coupon Buddy*. More specifically, the challenges in inputting coupon information reflect an area that is ripe for a data-labeling (comment classification) machine learning response, as discussed in the literature review (Sun’s *Label and Learn* system). For example, an auto-complete function for inputting coupon information based on previous user input could improve the process for labeling and organizing coupons, reducing the work required from the user.

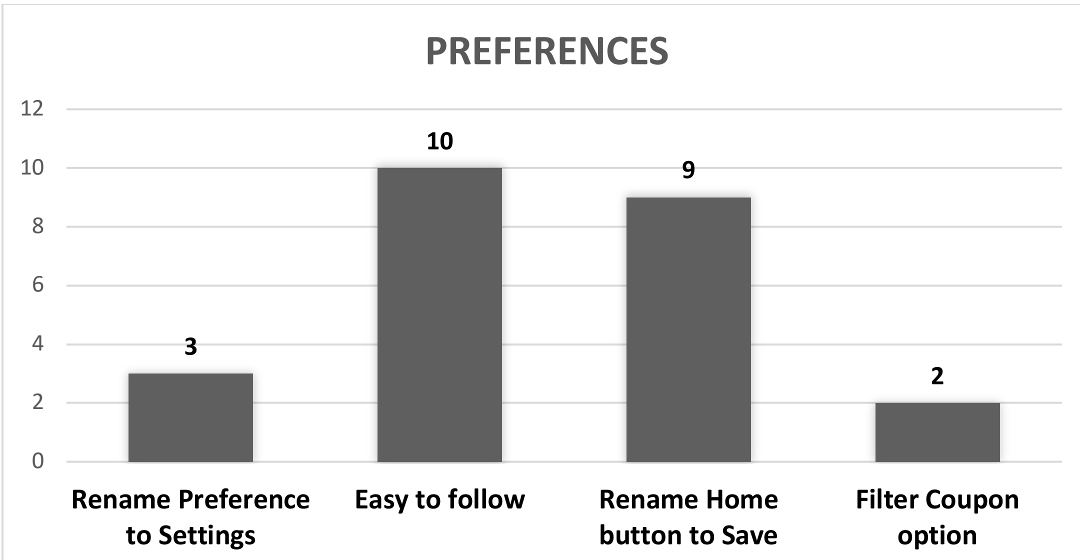

Figure 14: Preferences Prototype

According to the feedback, the information on the preferences page was largely self-explanatory and satisfactory for participants. Some suggestions *were* provided to guide us on additional features that could make the application more responsive with machine learning behavior. Two participants suggested adding a filter option to control what kind of coupons users would like to automatically add, even if there was an option for "look for similar coupons" once they added coupons. Although the app had the ability to look for similar coupons once users started inserting coupon details, participants mentioned users would need control to what kinds of "similar" coupons they would like to add.

One participant suggested "…Allowing users to clip these coupons themselves would give them additional control over their coupon list." There was also a debate over the language to label the buttons and titles in the preferences menu. Participants wondered if the language for "preferences" could be changed to "settings," which they believed is the most common term used by other applications (again, reflecting the power of app standardization). Finally, some participants also suggested renaming the button for "home" to "save settings," which they argued allows users to confirm that their changes have been accepted. Still, participants had no difficulty following the options available to adjust account settings.

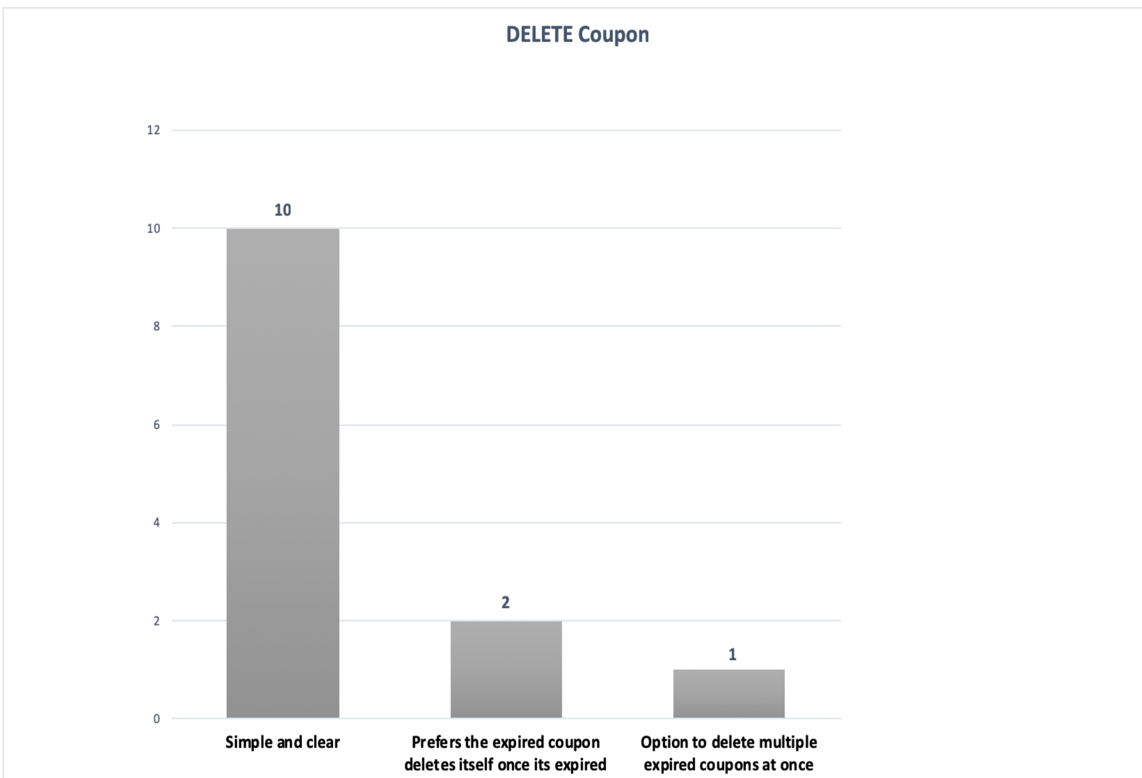

Figure 15: Delete Coupon Confirmation Prototype

Figure 15 demonstrates how participants proceeded to delete coupons manually once the coupons expired. Two participants believed that it was a waste of time to verify and manually delete expired coupons. One of the participants said "Why take the time to manage and delete? Just let it expire and fall off the list." On the other hand, other participants complimented the function by suggesting "Confirmation on delete is great." It appears some participants continue to feel it is crucial for users to know and confirm before deleting non-expired coupons. Other participants also asked about the ability to delete multiple expired coupons at once. These suggestions and recommendations on how *Coupon Buddy* should behave with expired coupons will be explored further when analyzing different usability behaviors; again, this field is ripe for machine learning application by using pattern recognition to tailor the desired deletion method to each user.

As we reviewed and analyzed how our participants commented on our prototypes with regard to potential design and functional improvements, we acknowledged that our participants were most concerned about design layout. They were intrigued about additional features that could be added to make the application integrate machine learning ideas. Our participants shared their feedback and thoughts for design improvements as if they were part of the research team; as if the application was on the cusp of actually happening. Our next and closing chapter will discuss the implications of all the feedback and its potential implementation in improving the app by applying comment classification and pattern recognition to integrate machine learning methods. We will assess the meaning of the results and speculate on what could be done to have UX designers become better designers by responsively applying machine learning algorithms within the classification and pattern recognition genres.

Chapter 4: Discussion

It may seem intriguing to employ machine learning improvements into an app by training user behavior interactions and improving learning data through user interactions. However, transforming systems into machine learning algorithms can be too complex to be explicitly programmed. Given the growing popularity of machine learning and artificial intelligence, we asked in the beginning of the thesis how we could imbibe the practice of machine learning into a designer's design thinking? Our research survey results in this chapter will summarize and discuss the main points from all the feedback and suggestions collected from our findings and elaborate on the lessons learned through this research study with regard to imbibing machine learning into design thinking. The feedback collected on coupon management (automatic deletion of expired coupons) and the addition of coupons is most lucrative for exploring machine learning classification and pattern recognition responses to user behavior. Machine learning algorithms in these two genres would tailor user experiences to manage how coupons are added, populated with information, sorted, and deleted.

In addition to tailoring sorting and deletion by imbibing initial user habits, machine learning in *Coupon Buddy* could implement versions of the data sorting (*Label-and-Learn*) and recommendation strategies described in the literature review. As the application evolves and builds a user base, UX designers could in turn "mine" the patterns and tentative classifications of user behavior to fine-tune machine learning algorithms to further improve classification and sorting techniques developed to respond to user behavior.

The Use of Three Different Prototypes

The primary reason behind designing three different versions for *Coupon Buddy* was to observe how participants would respond by reviewing different prototypes for our application. A number of conclusions can be drawn from the results: First, we realized that our participants' responses were correspondingly the same between different prototype versions. We had assumed that participants would respond to the incompleteness of the prototypes and to survey questions such as "What needs to be improved?" but we were surprised that responses were preoccupied with improvements, added functionalities, and machine learning potential rather than with the completeness or appearance of the mockup pages. Again, a majority of the feedback speculated about features that could be added to make the application adapt to a quality experience, and how it should respond to each unique user's behavior.

These surprising results may make it easier to improve *Coupon Buddy* and refine machine learning methods: since feedback for the prototypes was not appearance-based, it was easy to classify and sort comments to determine patterns in the feedback (i.e. less time wasted discarding irrelevant comments), which means the same approach may hold for imbibing machine learning into the application: any algorithms developed to refine application behavior to respond to specific patterns or classifications/commonalities of user behavior would likewise not need to be preoccupied (complexly-coded) to disregard irrelevant or trivial user interactions.

Major Findings

Although our study observed how detailed the suggestions were for the various *Coupon Buddy* mockups, we witnessed that some of our participants didn't always leave particularly useful or critical feedback that could be analyzed to improve our design or integrate machine learning concepts. As our participants were assigned to provide the best suggestions by observing eight different prototypes designed for each different prototype version, the length of feedback differed for each prototype. Some participants were acritical and possibly eager to move on when observing certain pages by leaving comments such as "Nice" or "Options are nice" or they even skipped some sections entirely (i.e. left no feedback).

This phenomenon was particularly pronounced on the feedback for the Adobe Illustrator prototypes, which looked the most polished or complete (even if in actuality the functionality behind the mockups was just as incomplete and ripe for improvement as for any of the other prototype versions). This strengthened our speculation that the *appearance of completeness or professionalism* subconsciously influences participants to less critically evaluate a prototype than if they evaluate a prototype that more explicitly looks and feels incomplete. The presentation of an excessively-polished mockup may dissuade surveyors from providing sufficiently critical feedback; the urge to conclude that a *polished appearance equals a polished idea* may be too strong. This has critical implication for machine learning methods: just as an overly-polished prototype may prompt participants to leave trivial or uncritical feedback, an overly polished and inflexible application may “script” diverse users to all behave the same way, hampering the potential for machine learning to customize the application and its functions to each user. In this kind of situation, pattern recognition and classification can be rendered moot if the entire sample (user base) is guided into behaving the same way.

This observation support the hypothesis that participants won’t compose lengthy feedback for the prototype if it isn’t interactive enough for them to test the application. Referring back to the literature review, Krause et al. stated “A challenge of crowdsourcing design feedback is that the results are often low quality. The reason for this may include diverse contributors who lack sufficient motivation, context, knowledge, and sensitivity to provide effective feedback (Krause et al., 2017).”

Responding to Major Findings for Future Prototypes

Now that we have speculated how participants responded to different versions of the prototypes, and how polished design appearances lead participants to provide less feedback for design improvements, we will vary our approach to create less polished prototypes and therefore elicit qualitative feedback data by emphasizing paper prototypes. This could in turn allow any machine learning methods applied to the application to better classify user patterns and respond to them. The most compelling evidence observed during the analysis was when we witnessed how participants responded to certain prompts by providing lengthy suggestions and recommendations for potential machine learning concepts. Not only would lengthy feedback benefit us in testing future design improvements but collecting as much feedback as possible would help UX designers mine user behavior patterns.

As we discussed in the literature review, UX designers struggle to work with data experts to integrate machine learning concepts due to low proficiency and knowledge of the concept. To circumvent this challenge and engage designers to become better, our future sessions will facilitate the design of additional rapid paper prototypes to collect sufficiently comprehensive data and deeply engage participants. The quality of feedback obtained in our database suggests that paper prototypes helped us receive feedback that was essential for design improvements. Not only did the feedback emphasize why paper prototypes generate better contributions and lengthy suggestions, it also helped us delve into further innovations essential to training and adapting machine learning experience.

To conclude the discussion about data analysis and summarize how our participants responded to differently designed prototypes, the varied responses allowed us speculate *why* suggestions differed based on prototype versions. Ultimately, our scope for this entire project was to analyze how the feedback, upon being classified via machine learning, could help UX designers take steps to integrate machine learning even further into design improvements. After all, design improvement is all about feedback, discussion, and observation. Conducting the research helped us acknowledge how collected feedback can be utilized to train machine learning interaction to respond to different suggestions made for different prototype versions. Finally, for future design improvements we will explore further why paper prototypes should be prioritized for implementing machine learning ideas.

Chapter 5: Conclusion

By conducting our research into how participants' respond to different prototypes, we discovered how best to implement machine learning methods with regard to classification of comments, feedback, and pattern recognition. Our research helped us ascertain how prototypes engaged our participants to contribute as many responses as possible and to potentially train the system for machine learning. Again, we should stress that *Coupon Buddy* was only a method and a task-driven *vehicle*; it merely facilitated and prompted participants' reactions to different designed prototypes. Although many of our participants were highly interested in using the application with machine learning functionalities, our ultimate approach was to observe how we could apply machine learning to the classification of responses and imbibe that into design thinking.

For example, recalling our survey analysis chapter, our most illuminating example of integrating machine learning into the application by classifying and responding to the most common feedback was the most frequently-discussed suggestion of *having the ability for camera accessibility to take pictures while adding a coupon.*

Introducing changes to the application by adding machine learning features would make it more effective in both conveying and adapting to the purpose of the application's existence and premise. Although much of our feedback discussed machine learning experiences and design improvements, all the observations collected into the database were ideal to begin training the system to adapt to user behavior. Given the 83 participants involved, we realized that training the system to respond to suggested ideas by analyzing patterns and classifying responses and interactions can make the application significantly more intuitive and useful.

We were fascinated to learn that our participants already thought like user experience designers to provide feedback that could be used to train a machine learning classification system, and for us to provide less polished information for the purpose of prototyping. Given how comfortable participants are in evaluating prototypes – even knowing much of the terminology and best practices of user experience design – and also given how our most productive feedback came from the paper prototypes (in which the most useful feedback emerged from a level of incompleteness in the appearance of the prototypes to prompt users' imaginations, in contrast to the highly-polished Illustrator prototypes), we conclude with a combination recommendation and question:

It is typical for application developers to provide fairly polished prototypes for surveying, alpha/beta testing, and general user feedback, but given our experience with rough paper prototyping being the most fruitful prototype genre for productive user feedback, might it make sense for other app designers to pull back from providing typically-polished prototypes and mockups for feedback in favor of rougher, even more abstract layouts to garner more critical feedback? Our research results suggest that something about the "blank slate" approach spurs an imagination from participants and prompts them to be more critical in their analysis – even to look for opportunities to exploit and add improvements that the prototype designers did not initially think of – whereas the more polished prototypes seemed to elicit a more passive, less useful approach: *everything looks great and polished so I'll just breeze past this*.

The implications for machine learning development, testing, and feedback accumulation could be tremendous: not only might the quality and breadth of the feedback be better with paradoxically "rougher' prototypes, but the development of said rougher prototypes would save considerable time in the early phases of app development (where feedback is more crucial), pushing the polishing of the user interface into the later stages where the collected feedback will be better situated to inform the final design and integrate machine learning approaches.

References

Chiu, C.-M., Liang, T.-P., & Turban, E. (2014). What can crowdsourcing do for decision support? *Decision Support Systems*, *65*, 40–49. https://doi.org/10.1016/j.dss.2014.05.010

Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17 (pp. 278–288). Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025739

Hamid, S. S. A., Admodisastro, N., & Ghani, A. A. A. (2015). Computer-based learning model to improve learning of the malay language amongst dyslexic primary school students. In *Proceedings of the Asia Pacific HCI and UX Design Symposium on - APCHIUX '15* (pp. 37–41). Melbourne, Victoria, Australia: ACM Press. https://doi.org/10.1145/2846439.2846448

Krause, M., Garncarz, T., Song, J., Gerber, E. M., Bailey, B. P., & Dow, S. P. (2017). Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (pp. 4627–4639). Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025883

Lee, T. Q., Park, Y., & Park, Y. T. (2008). A time-based approach to effective recommender systems using implicit feedback. Expert systems with applications, 34(4), 3055-3062.

McDuff, D., Karlson, A., Kapoor, A., Roseway, A., & Czerwinski, M. (2012). AffectAura: an intelligent system for emotional memory. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (p. 849). Austin, Texas, USA: ACM Press. https://doi.org/10.1145/2207676.2208525

Stumpf, S., Sullivan, E., Fitzhenry, E., Oberst, I., Wong, W.-K., & Burnett, M. (2008). Integrating rich user feedback into intelligent user interfaces. In *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08* (p. 50). Gran Canaria, Spain: ACM Press. https://doi.org/10.1145/1378773.1378781

Sun, Y., Lank, E., & Terry, M. (2017). Label-and-Learn: Visualizing the Likelihood of Machine Learning Classifier's Success During Data Labeling. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17* (pp. 523–534). Limassol, Cyprus: ACM Press. https://doi.org/10.1145/3025171.3025208

Tsukada, K., Mizushima, Y., Ogata, A., & Siio, I. (2010). LetterTwitter: smart mailbox for spam-filtered notification of received letters. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Ubicomp '10* (p. 439). Copenhagen, Denmark: ACM Press. https://doi.org/10.1145/1864431.1864476

Yang, Q., Banovic, N., & Zimmerman, J. (2018). Mapping Machine Learning Advances from HCI Research to Reveal Starting Places for Design Innovation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (pp. 1–11). Montreal QC, Canada: ACM Press. https://doi.org/10.1145/3173574.3173704

Yang, Q., Scuito, A., Zimmerman, J., Forlizzi, J., & Steinfeld, A. (2018). Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18* (pp. 585–596). Hong Kong, China: ACM Press. https://doi.org/10.1145/3196709.3196730

Yang, Q., Suh, J., Chen, N.-C., & Ramos, G. (2018). Grounding Interactive Machine Learning Tool Design in How Non-Experts Actually Build Models. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18* (pp. 573–584). Hong Kong, China: ACM Press. https://doi.org/10.1145/3196709.3196729

Yang, Q., Zimmerman, J., Steinfeld, A., & Tomasic, A. (2016). Planning Adaptive Mobile Experiences When Wireframing, 13.

Appendix A: Consent Form

<u>Whom to Contact about this study:</u>
Principal Investigator: Tauhid Ahmed, Student, Interaction Design and Information Architecture
Department: Yale Gordon College of Arts and Science
Telephone number: 443-629-3862

CONSENT FORM FOR PARTICIPATION IN RESEARCH ACTIVITIES

I. <u>INTRODUCTION/PURPOSE:</u>

I am being asked to participate in a research study to help the researcher provide my feedback and recommendations for his mockup application. The feedback data will help the researcher analyze the comments, and adapt machines learning practices to learn and express user behavior for the application.

II. <u>PROCEDURES:</u>

As a participant in this study, I will be assigned to look through several design prototypes and leave best suggestions and comments for each designed mockup through a web-based survey. Each of my responses will be stored in a password protected database, which will be used by the researcher to teach the application machine learning algorithms. The feedback given under each screen will help the researcher explore strategic analysis on the design in terms of intuition, language of use, label coordination, and if the design "looks and feels" complete.

This should take about 10-15 minutes.

III. <u>RISKS AND BENEFITS:</u>

My participation in this study does not involve any significant risks and I have been informed that my participation in this research will not benefit me personally, but the results or outcome of study will benefit user experience designers understand better by adapting the practice of machine learning in design thinking to reduce redundancy and static designing, and to cultivate them to become better designers.

IV. <u>CONFIDENTIALITY:</u>

Any information learned and collected from this study in which I might be identified will remain confidential and will be disclosed ONLY if I give permission. All information collected in this study will be stored in a database system. Only the investigator and members of the research team will have access to these records. If information learned from this study is published, I will not be identified by name. By clicking accept at the bottom of this form, however, I allow the research study investigator to make my records available to the University of Baltimore Institutional Review Board (IRB) and regulatory agencies as required to do so by law.

Consenting to participate in this research also indicates my agreement that all information collected from me individually may be used by current and future researchers in such a fashion that my personal identity will be protected. Such use will include sharing anonymous information with other researchers for checking the accuracy of study findings and for future approved research that has the potential for improving human knowledge.

V. CONTACTS AND QUESTIONS:

The principal investigator, Greg Walsh, has offered to and has answered any and all questions regarding my participation in this research study. If I have any further questions, I can contact Dr. Greg Walsh at (gwalsh@ubalt.edu) or 410-837-5473 or I can contact the University of Baltimore's Institution Review Board Coordinator at (irb@ubalt.edu) or 410-837-6199

VI. VOLUNTARY PARTICIPATION

I have been informed that my participation in this research study is voluntary and that I am free to withdraw or discontinue participation at any time.