Empowering Learners with a Low-Barrier Mobile Data Science Toolkit

Mobile Data Science Toolkit

Using MIT App Inventor to build data science mobile applications.

Anonymous*

Anonymous affiliation

This paper introduces a novel data science toolkit designed specifically for children, enabling them to effortlessly create mobile apps integrated with data science capabilities. The toolkit showcases new features that simplify the data science process for young users. Additionally, the paper presents a collection of example apps created using the toolkit, highlighting the versatility and potential of this innovative platform. By empowering children to explore data science through app development, this toolkit opens exciting opportunities for hands-on learning and creative expression in the field of citizen science.

Keywords and Phrases: MIT App Inventor, Data Science, Mobile Application

1 INTRODUCTION

Many barriers, such as cost, access, and complexity, can prevent high school students from building working prototypes for performing authentic data science practices across the data science life cycle — collecting data, data cleanup, data visualization, and prediction [1]. For example, online tools such as Python notebooks can be too complex for students new to Python, and such tools only allow users to perform parts of the data science life cycle. Moreover, few products allow students to build projects that visualize data on mobile platforms — which, after all, is where many scientists and citizen scientists collect and consume data visualizations today [2].

We discuss and present demos and student examples for a low-barrier mobile toolkit that allows learners to take part in the whole data science lifecycle. The demo and discussion are relevant for teachers in middle school and high school, and facilitators of out-of-school programs. Participants can build simple mobile apps using MIT App Inventor and test these systems for themselves during the conference presentation.

The toolkit is a free suite of new, block-based-programming data science features that allow students to solve real-world problems that affect them or their community. In the following paragraphs, we describe important features we have implemented. The next section presents example apps created by such features.

* Anonymous

1.1 IOT Connection and Data Management Features for Mobile Phone Apps

The new App Inventor IoT connection component focuses on establishing a seamless connection between the mobile phone app and a low-cost Bluetooth sensor. This connection enables real-time data collection from the sensor, ensuring accurate and reliable time series data acquisition. By leveraging Bluetooth technology, the app can connect to the sensor wirelessly, eliminating the need for complex wiring and enhancing the mobility of the data collection process [3, 4]. The online spreadsheet to mobile device component streamlines the data from spreadsheets to the mobile device. This functionality allows users to visualize, clean, and analyze the collected time series data conveniently. By integrating online spreadsheet services, such as Google Sheets or Microsoft Excel Online, the app facilitates easy access to data from anywhere, fostering collaborative data analysis and reducing the dependency on traditional desktop-based software. Additionally, cloud storage services can be integrated into the app, enabling seamless synchronization and backup of data to remote servers.

1.2 Anomaly Detection and Data Cleaning Techniques for Mobile Phone Data Analysis

The Anomaly Detection component focuses on identifying anomalies and deciding whether to remove them based on the type of data and the user's domain knowledge. By incorporating advanced anomaly detection algorithms, the application can automatically flag potential anomalies in the data. However, it recognizes that not all anomalies are erroneous and may reveal valuable insights depending on the specific domain. Therefore, the feature allows users to exercise their domain knowledge and make informed decisions on whether to remove flagged anomalies or retain them for further analysis [5]. The Preliminary Visualization feature enables users to perform initial visual analysis on their mobile phone's touchscreen. By providing interactive visualization tools, the application facilitates the identification and removal of out-of-context anomalies. Users can explore the data visually, observing patterns, trends, and irregularities directly on their mobile device. This feature empowers users to make data-driven decisions on removing anomalies that are visually distinguishable and do not align with the expected context of the data.

1.3 Mobile Charting for Visualizing Data in Mobile Applications

The Mobile Charting feature focuses on providing users with multiple options for visualizing their data in a mobile application. By incorporating chart types such as bar graphs, line graphs, scatter plots, and pie charts, users can select the most suitable representation for their data based on the nature and characteristics of the dataset. This flexibility allows for effective communication and comprehension of the data, enabling users to identify patterns, trends, and relationships.

1.4 Linear Analysis for Predictive Insights in Mobile Applications

The Linear Analysis feature focuses on providing users with a line of best fit to showcase trends in the data. By fitting a straight line to the data points, users can observe the direction and magnitude of changes, helping them anticipate future outcomes. This feature enables users to understand the relationship between variables and make predictions based on the linear trend. The Linear Analysis feature also includes presenting statistical and probability information about the data. This information aids users in assessing the strength and significance of the linear fit. For example, the feature provides correlation coefficients to quantify the degree of linear association between variables. These statistical insights enhance users' understanding of the data and assist them in making informed decisions.

2 EXAMPLES

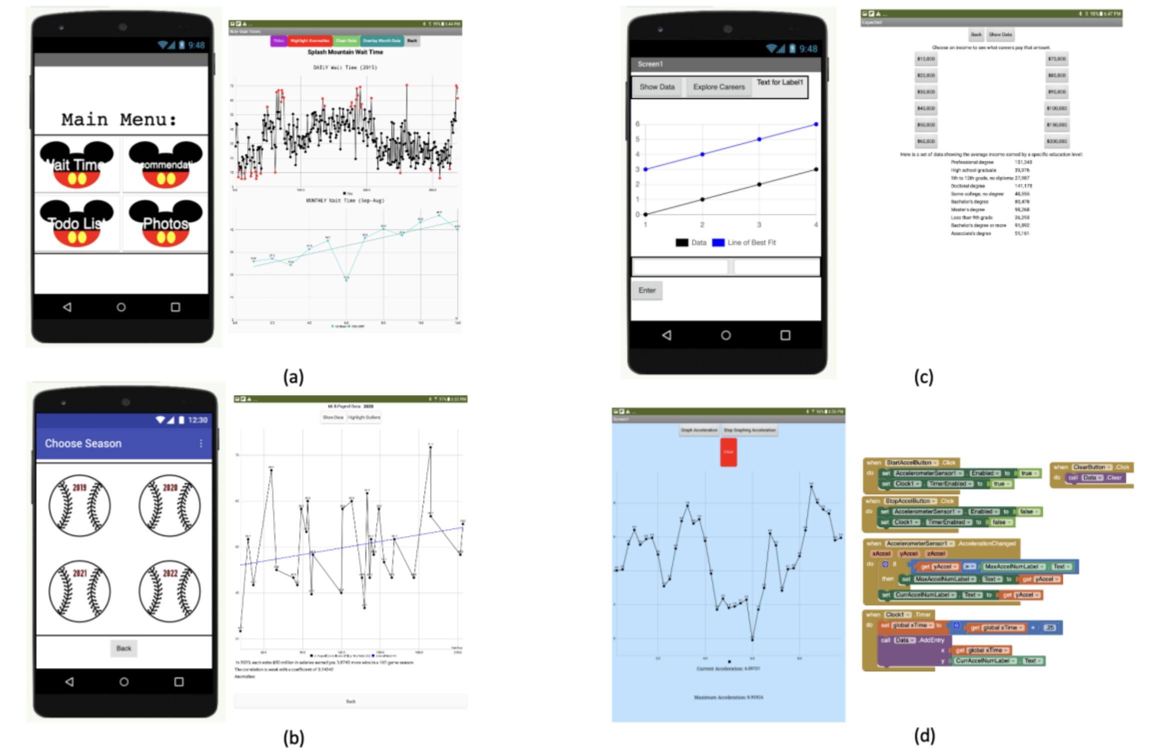

Figure 1: Images of sample apps using data science toolkit.

2.1 Analyzing Posted Wait Times for Disney World Rides: Insights for Average Wait Time Prediction and Park Attendance Evaluation

Figure 1 (a) focuses on analyzing the wait times for various rides at Disney World to determine the average time a child might have to wait in line. Raw data obtained from a reliable source, specifically the Touring Plans website provides information on ride wait times throughout the year. It is based on a scenario where children want to estimate their average wait time for a ride at Disney World. The data is graphed to identify trends in ride busyness during specific times of the day, month, or year. Additionally, the app explores the significance of extremely low and high wait times, allowing children to understand factors such as weather conditions, park attendance, or ride maintenance that contribute to these outliers.

2.2 Analyzing the Relationship Between Payroll and Wins in Major League Baseball: An Anomaly Detection Approach

Figure 1 (b) focuses on investigating the relationship between team payroll and wins in Major League Baseball (MLB). It highlights the significance of payroll expenditure compared to the number of wins achieved by each team. With over 150 games played by each team in a season, the app analyzes how spending on player salaries impacts a team's performance. It also acknowledges the presence of outliers, such as teams that overspend or underspend on player salaries, as exemplified in the movie "Moneyball." Consequently, anomaly detection techniques are employed to identify and analyze such outliers, making this example a suitable choice for anomaly detection in the context of MLB payroll and wins.

2.3 Exploring the Relationship Between Education Level and Average Income: Anomaly Detection and Trend Analysis

Figure 1 (c) investigates the relationship between education level and average income using a raw dataset. It focuses on a unique anomaly wherein the traditional trend of higher education level leading to higher annual income is disrupted by the doctoral degree. Instead of conforming to the expected pattern, the anomaly reveals a shift in the trend. This example challenges children to critically think about the potential reasons behind this unexpected observation. This app employs anomaly detection techniques and trend analysis to uncover and analyze this intriguing anomaly, encouraging children to explore the complexities of the relationship between education and income.

2.4 Exploring Accelerometer Applications in Sports: From Vertical Jump Height to Sprint Performance Analysis

Figure 1 (d) delves into the diverse applications of accelerometers in the field of sports. It focuses on several potential uses of accelerometers, including measuring vertical jump height, evaluating push-up performance, assessing throwing speed, investigating the impact of different factors on sprinting, and tracking hydration levels for sprint analysis. Additionally, it explores the possibility of incorporating weather data in conjunction with accelerometer measurements. By engaging children in hands-on activities and data tracking, this research aims to foster interest and understanding of the application of accelerometers in sports performance analysis.

3 CONCLUSION

By providing an accessible data science toolkit, this paper enables students to actively engage as data scientists in areas aligned with their interests and experiences. The inclusion of new features empowers young users to explore data-driven concepts, analyze real-world data, and create meaningful mobile apps that reflect their unique perspectives. This approach not only fosters a deeper understanding of data science but also encourages students to apply their newfound skills and knowledge in diverse domains, promoting a sense of ownership and relevance in their data science journey.

ACKNOWLEDGMENTS

Anonymous

REFERENCES

[1] Mike, K., Ragonis, N., Rosenberg-Kima, R. B., & Hazzan, O. (2022). Computational thinking in the era of data science. Communications of the ACM, 65(8), 33-35.

[2] Silvertown, J. (2009). A new dawn for citizen science. Trends in ecology & evolution, 24(9), 467-471.

[3] Lechelt, S., Rogers, Y., & Marquardt, N. (2020, June). Coming to your senses: promoting critical thinking about sensors through playful interaction in classrooms. In Proceedings of the Interaction Design and Children Conference (pp. 11-22).

[4] Clark, J., Falkner, W., Balaji Kuruvadi, S., Bruce, D., Zummo, W., & Yelamarthi, K. (2019, March). Development and Implementation of Real-Time Wireless Sensor Networks for Data Literacy Education. In Proceedings of the 2019 ASEE North Central Section Conference, Morgan Town, WV, USA (pp. 22-23).

[5] Vaidya, A., & Sharma, S. (2023). Anomaly detection in the course evaluation process: a learning analytics–based approach. Interactive Technology and Smart Education.