TruMark: Attendance Recording Website with Machine Learning Verification

-

Cuneyt Aksoy

001385752

ca4146@truman.edu

Truman State University

B.S., Computer Science

Senior Capstone Project Report

Supervised by Dr. Kafi Rahman, Ph.D.

Fall 2023

November 26, 2023

Table of Contents

1. Abstract

In the rapidly evolving realm of education, there is an increasing demand for effective, technology-driven solutions that can streamline administrative tasks and elevate the overall academic experience. As traditional modes of teaching and learning continue to be transformed by innovation, universities require more intuitive systems that simplify processes, ensure security, and enrich learning. The proposed project, TruMark, directly addresses this growing need within the Truman State University community by introducing an inventive, web-based attendance tracking system uniquely tailored for Truman students and faculty.

Leveraging machine learning algorithms and signature analysis for user verification, TruMark aims to enhance and modernize attendance tracking on campus. Students can seamlessly log their attendance for courses through an intuitive interface. By automating parts of the attendance tracking process, TruMark provides convenience and helps ensure data integrity. The system is designed wholly with the Truman end-user experience in mind, from its user-friendly platform to its ability to monitor course attendance.

Drawing inspiration from existing University platforms like TruView and BrightSpace, TruMark looks to fill a void in a practical everyday need in the academic setting. By streamlining the attendance tracking process, TruMark seeks to enrich the academic experience at Truman State University, serving as a stepping stone on the path toward a more intuitive, technology-driven environment.

Keywords: TruMark, attendance tracking, machine learning, signature analysis, educational technology, user verification

2. Introduction

2.1 Background

The landscape of teaching and learning has undergone immense changes in recent decades, propelled by rapid technological advancement. Where classrooms were traditionally confined by physical space and timed class periods, innovation now provides opportunities to improve the flexible, engaging academic experience. As learning moves beyond four walls into blended and online methods, administrative systems must also adapt to support evolving pedagogies and regulatory requirements.

Routine tasks like attendance tracking that previously required manual effort now require intuitive, data-driven digital solutions. The emerging potential of educational technologies to streamline back-end processes while elevating front-end instruction has driven increasing interest from all involved parties in academic institutions. Students arrive on campuses expecting seamless, mobile-centered services, while administrators seek to maximize institutional efficiency and compliance. At the same time, instructors face growing pressure to monitor student performance and engagement in larger classroom sizes and teaching loads.

For faculty in particular, attendance marking presents a tiresome cycle of paperwork across courses each term. Traditionally relying on paper forms or call-and-response methods, the process distracts from pedagogy and wastes valuable minutes of every class period. Further, manual records pose data management issues, making it difficult to track trends, identify struggling students proactively, or produce specialized reports on demand. A centralized electronic system could save time through simple login verification from any device, while automatically compiling accurate data for instructors and advisors alike.

Students would also benefit tremendously from streamlined attendance tracking. Current interruptions to call roll disrupt learning flow and diminish limited class time that could be spent on dynamic instruction. Younger digital natives expect user-friendly, mobile-first services as the norm across life domains. A technology-driven check-in process would empower independent responsibility for attendance without comprising pedagogical minutes, potentially boosting class engagement and performance indicators over the long run.

From an administrative perspective, higher education leaders acknowledge strategic and budgetary value in combining disparate systems onto controlled, secure university platforms. Paper-based methods lack oversight and introduce risks of loss, theft or error compared to centralized digital records. Automated reporting functionalities are also necessary to adhere to changing regulations and assess impacts of interventions, such as providing guidance to students flagged for lack of engagement or substantive interaction according to university policy.

While learning management platforms at institutions partially address these needs, attendance tracking often remains a manual outlier. At Truman State University specifically, successful systems like TruView and Brightspace currently focus on course materials, communications and grading rather than participation and attendance records. Given common practice elsewhere and past experiences, it is reasonable to expect that some Truman faculty may continue relying on physical documentation of daily attendance or informal technology solutions as alternatives, at least in the interim. A custom-designed, university-hosted solution could systematically refine this process by streamlining workflows across instructive and operational fields to maximize impact for all campus stakeholders.

2.2 Problem Statement

Current attendance tracking methods at Truman State University present several challenges that TruMark aims to shed light on. Reliance on paper records introduces significant inefficiencies and vulnerabilities that undermine the integrity of important student data over time. Physical documents are redundant, prone to loss or damage, and incompatible with distributed learning environments increasingly prevalent post-pandemic. More concerningly, manual processes preclude integration of participation records with other crucial systems, missing opportunities to leverage full-spectrum analytics promoting student success.

Paperwork requirements have the possibility to impose immense, repetitive costs on faculty who must dedicate precious minutes intended for dynamic instructional activities to redundant filling of forms across multiple courses instead. This wasted effort significantly takes away time available for higher-impact learning experiences and teaching innovation (Kim, 2020). Automating the process through streamlined electronic workflows could free hundreds of faculty hours annually for strategic reallocation.

The rapid transition to online education during the COVID-19 pandemic worsened these issues and inefficiencies. As classes shifted online, the average amount of time it took to take attendance was much greater, averaging from five to seven minutes compared to only one to two minutes for in-person classes (Kim, 2020). With students and faculty adjusting to distance learning formats, administrative tasks like attendance tracking placed an even greater time burden on instructors. Streamlining electronic attendance taking processes could help address inefficiencies that became more pronounced with expanded online and hybrid teaching models. Reducing time spent on redundant documentation would allow a strategic reallocation of minutes to more impactful pedagogical activities, ultimately benefitting student learning outcomes in classrooms.

Without centralized digitization and validation protocols, inaccuracies naturally creep into handwritten records over time, threatening the integrity of important datasets used for compliance reporting, advising interventions, and administrative decision-making. Manual methods also introduce security vulnerabilities compared to secure digital repositories governing access and utilization of sensitive student data. As learning shifts increasingly online amid growing student expectations for optimized experiences, piecemeal approaches lack cohesion and usability.

Fundamentally, the lack of an integrated electronic system designed specifically for Truman's academic community represents a missed opportunity to leverage comprehensive analytics empowering continuous advancement. Beyond core attendance collection functionality, opportunities exist to integrate participation insights with other crucially intersecting data sources like grades, early alerts, program assessments, and more to generate actionable intelligence supporting evolution toward fully technology-driven strategic operations optimized for university stakeholders.

2.3 Objectives

The overarching goals of the TruMark system are to simplify attendance tracking workflows, ensure accurate attendance data, provide convenient accessibility, enhance the user experience, and most importantly, serve as a stepping stone toward further integrating technology into the academic experience at Truman State University, especially in the context of attendance tracking.

The first objective is to simplify attendance tracking through a centralized web-based platform. Currently, the reliance on paper-based methods has proven inefficient and time-consuming for both students and faculty. Manually marking attendance disrupts the flow of every class period. TruMark aims to streamline this process by allowing users to easily log their attendance online from any device with a web browser. With a simple login at the start of class, time can be preserved for dynamic learning rather than administrative tasks. For students, TruMark provides a user-friendly way to verify their presence that is less interrupting than calling roll. Independence in tracking attendance promotes responsibility while freeing faculty from paperwork. They can instead focus on delivering engaging instruction. Time savings at scale across courses could recoup hundreds of faculty hours annually.

The second goal of ensuring accurate attendance records requires validating user identities to protect privacy. TruMark incorporates machine learning algorithms and signature analysis during login to authenticate the user. This level of access control and data security strengthens integrity compared to manually-completed paper forms vulnerable to errors and copying.

While paper-based methods remain common, the limitations of this approach might lead some faculty to resort to other workaround options for attendance recording. In particular, a number of third-party software platforms and apps exist that allow attendance to be tracked digitally. However, these proprietary systems pose several security and privacy issues when utilized without centralized governance. Records maintained on external systems are beyond the university's direct control and oversight. Sensitive student data could be at risk if third-party vendors do not uphold the rigorous security standards expected by Truman State University and protect confidential information. Data breaches of external platforms not customized for the organization could unintentionally compromise student privacy in several ways. Proprietary third-party systems are designed primarily to meet the needs of their general user base, rather than fulfilling the unique privacy, security and compliance requirements of an academic institution. They may not implement stringent access controls, authentication protocols, or data encryption practices required to safeguard sensitive educational records.

If a breach occurs, it could potentially expose personally identifiable information like student names, IDs and attendance records. This places Truman State University at legal and reputational risk if student privacy is not sufficiently protected according to stringent federal standards for educational data. Compromised records could also enable malicious actors to exploit stolen student identities for fraud or social engineering. From the perspective of students and parents, a privacy incident may damage trust in the university's ability to be a responsible steward of personal data.

Having attendance and participation records stored externally on third-party platforms also means Truman has less visibility and control over how data is collected, used and potentially shared with other entities through the vendor's own business operations. The university cannot thoroughly oversight privacy practices it does not directly manage. This decentralized approach limits Truman's autonomy and governance over critical student information assets.

Additionally, records maintained through diverse external systems become fragmented and difficult for faculty themselves to consistently access or aggregate insight across changing courses and semesters. There is no organized way for the university to monitor participation trends, identify at-risk students, or generate whole-of-institution analytical reports from disparate sources. Differing platforms also introduce non-uniformity in methodology that undermines consistency relied upon for auditing and compliance purposes. A single centralized system designed specifically with the needs of Truman's academic community in mind helps ensure streamlined, secure processes and optimized experiences for all stakeholders.

By providing a trusted online solution fully governed by the university, TruMark aims to eliminate these issues by doing away with the piecemeal use of external tools to centrally house accurate attendance data within the organization's controlled infrastructure. Additionally, by automating the attendance tracking process, TruMark stands to directly support Truman's Substantive Interaction policy which aims to ensure meaningful engagement is occurring regularly between faculty and students. Accurately recording participation through digital check-ins simplifies documenting substantive interactions as defined by federal regulations.

Beyond the system, accurate longitudinal records empower more robust reporting to assess compliance, interventional impacts and inform strategic planning. Currently, siloed paper archives scattered across departments provide limited visibility. TruMark aims to centralize once-disparate sources of attendance insights into connected datasets.

While modest in current scope, the vision of TruMark lies in strategically streamlining Truman's academic processes through user-centered technology. Automating attendance collection represents an initial step toward a more integrated, data-driven support framework empowering all campus stakeholders. Eventually, insights could inspire innovative new pedagogical models maximizing engagement in any learning environment. Ultimately, Truman State University's goals around meaningful student-faculty contact are directly supported through reforming attendance tracking workflows with an optimized digital solution.

2.4 Significance & Scope

While the initial TruMark application focuses on automating the basic process of attendance logging through digital check-ins, its greater purpose is to establish meaningful groundwork for the future of online education at Truman State University. As the capabilities of educational technology continue to rapidly progress, entirely new models for monitoring participation in online or hybrid learning environments may become feasible.

Through serving as an early foray into centralizing attendance tracking via an online system, TruMark provides the opportunity for Truman to begin exploring a variety of potential future capabilities to supplement or enhance attendance verification. Advances in areas like collaborative work tracking, learning analytics, and real-time engagement monitoring could eventually allow attendance to be verified through direct observation and assessment of student participation within virtual class sessions.

For example, presence may one day be automatically tracked through minutes spent actively contributing to online group assignments, projects or discussions based on detailed activity logs and time stamps. Engagement detection technologies may also allow things like monitoring keystroke frequency, scrolling and reading behaviors, or facial expressions for signs of passive viewing versus active involvement in online class meetings through cameras.

As learning management platforms integrate expanded analytics capabilities, indicators of participation like comments, questions, and file access trails could provide a more holistic view of attendance beyond just logins. TruMark helps establish a centralized foundation for aggregating diverse engagement data sources as they emerge. Ultimately, evolving technologies may open new avenues to gauge attendance beyond merely logging in, instead authentically verifying cognitive attendance through metrics of applied learning actions.

Access to enriched participation metrics linked across courses becomes invaluable for proactive advising. Timely analysis of attendance relationships to other metrics may reveal unseen impact areas to guide incremental process improvements. For example, triggering early alerts for at-risk students per attendance drops or substantiating policy decisions with analytics.

By serving as Truman's initial foray into streamlining attendance through online systems, TruMark therefore strives to lay the strategic groundwork to inform potential pilots of cutting-edge attendance solutions as innovation enables accurate new methods of monitoring presence in online or blended environments. Overall, while in early stages today, TruMark establishes an important foundational building block. Centralizing electronic attendance data collection standardizes analytical opportunities and positions the university to explore innovative verification methods as educational technologies continue advancing. The system primes Truman for strategic evaluations and pilots going forward.

3. Literature Review

3.1 Impact of Attendance Tracking in Higher Education

A growing body of research has begun to evaluate the impacts of online and automated attendance tracking systems compared to traditional paper-based methods. As educational technologies have advanced, several studies have assessed the relationships between digital participation monitoring and key student outcomes at the higher education level. While the primary focus of this literature review will be on attendance tracking within universities and colleges, some sources may also include discussions of K-12 schools to provide a more well-rounded perspective.

Deducing direct relationships between digital attendance systems and academic performance has proven challenging due to the myriad interconnecting influences on student success. Isolating the discrete impact of variables like online participation tracking is complicated by an intricate web of individual, environmental, situational, and circumstantial factors simultaneously shaping outcomes. Personal attributes such as cognitive abilities, socioeconomic background, earlier educational experiences, and psychosocial traits inherently differ between students and feed into developing trajectories. Contextual settings including instructional design, class dynamics, course characteristics, and available resources vary significantly across learning institutions as well. Even life events outside the academic setting can disrupt progress if major challenges arise.

Moreover, any engagement itself may simply indicate underlying qualities fostering perseverance rather than causally driving retention completely through involvement alone. Those regularly attending may innately exhibit stronger determination independent of actual attendance effects. Investigations try to partial out these elements statistically and experimentally. However, the dense interrelationships within educational systems defy clean disassociation of variables in isolation. Natural learning milieus house intricacies exceeding controlled scientific parameters.

While precluding absolute attribution, the collection of evidence still provides strategic insight. Recurring themes highlight technology's untapped potential to optimize processes, oversight, intervention timing, and experiences when tailored judiciously. This review explores meaningful takeaways amid research complexities intrinsic to the field. The aim is informed guidance, not oversimplified conclusions.

Additionally, the scope of research in this area examines far more than can be thoroughly detailed within the bounds of this capstone project. Sources consider numerous educational, social, and psychological dimensions that fall outside the direct objectives of automating Truman's attendance processes. For conciseness and relevancy, discussion of various theories and findings will focus only on the most pertinent points relative to benefits and usage of centralized digital solutions.

The following section aims to set the stage for sources reviewed by elaborating on emerging insights and themes within this field. Key areas of research assessing links between participation tracking technologies and metrics such as student satisfaction, retention and academic performance will be surveyed at a higher level. This overview establishes the strategic rationale and need for reforming outdated paper-reliant methods through user-focused innovation.

The paper "*Online engagement and performance on formative assessments mediate the relationship between attendance and course performance*" by Lu and Cutumisu examines an important issue in modern higher education - how the rise of technology-enhanced learning has impacted the relationship between class attendance and student performance.

As the authors note, previous research on this topic was conducted primarily in the context of traditional face-to-face courses. However, the growing availability of online resources has prompted a debate around whether regular in-person class attendance remains as vital for academic success in today's digital learning environments (Lu & Cutumisu, 2022).

To help address this debate, Lu and Cutumisu investigate how factors like online engagement and performance on formative assessments interact with the relationship between class attendance and end-of-course grades. Their study analyzes log file data extracted from the learning management system of a university course, allowing objective measurement of students' in-class attendance patterns as well as their online activities and quiz scores.

Through structural equation modeling techniques, the researchers find that while attendance alone does not directly predict grades, it positively impacts performance indirectly by influencing levels of self-regulated online learning and formative assessment scores. This suggests technology may compensate for lower attendance by providing alternate avenues for engagement.

Real-time data collection in education is becoming increasingly important as it allows for the tracking of program activities and the progress of instructors and learners. This data can inform adaptive management in education, enabling policymakers, practitioners, and researchers to make data-informed decisions and ensure access to quality teaching and learning experiences. Digital tools offer several advantages for real-time data collection compared to paper-based systems. They provide more efficient collection and analysis of real-time data, and offer flexibility and customizability. Digital technologies can also generate visualizations and recommendations automatically.

Real-time performance data can play a crucial role in addressing the global learning crisis by allowing for adaptations to program inputs, activities, and outputs. With real-time data, administrators, instructors, students, and families can tailor programs, ensure teaching at the right level, track progress, and advocate for needs. However, collecting data in real time poses challenges, as many countries lack sufficient data and/or the capacity to measure and monitor learning outcomes. The COVID-19 pandemic has exacerbated the learning crisis, with marginalized populations being particularly affected. Robust real-time data is crucial for informing policymaking, tracking progress, and ensuring accountability among stakeholders (Gustafsson-Wright, Osborne, & Aggarwal, 2022).

The use of digital tools for real-time data collection in education is significant for policymaking and research efforts to strengthen education systems. Real-time data can support the development of new tools, selection from existing tools, and data-informed decisions to ensure quality teaching and learning experiences for all (Gustafsson-Wright, Osborne, & Aggarwal, 2022).

The research study conducted by Bowen, Price, Lloyd, and Thomas at the University of Glamorgan in Wales provides important insights into different attendance monitoring approaches in higher education. Attendance tracking has become a critical focus for universities as research shows student attendance correlates with academic success and retention. However, traditional paper-based attendance tracking methods have posed numerous problems that limit their effectiveness (Bowen et al., 2005).

To address these challenges, the researchers developed and piloted an electronic attendance monitoring system called Uni-Nanny at their university. They compared this new automated method against the traditional manual paper-based system that the university had relied on for years. The paper-based system required professors to pass around sign-in sheets in each class and then manually enter and analyze the collected attendance data. This resulted in many limitations: inconsistency in procedures across different courses, missing or incomplete data, delays in accessing information, and the ability for students to easily falsify attendance records for absent classmates. The burdensome process of manual data entry and aggregation also constrained systematic analysis of attendance trends. The decentralized nature of the paper records also meant delays in staff accessing and consolidating the attendance information across campus. The burdensome nature of manual data entry and aggregation inhibited large-scale or university-wide analysis of attendance patterns and academic issues.

In contrast, the Uni-Nanny system automated the attendance tracking process through campus-wide technology integration. The researchers set up electronic attendance terminals across university buildings. Students would simply swipe their ID cards at the terminals to register their attendance for each class. This instantly captured student attendance data and wirelessly transferred it to a centralized database in real-time. Unlike the paper records, the attendance data was complete, consistent, and easily consolidated across courses. With just a few clicks, professors and authorized staff could securely access up-to-date attendance reports on both current classes and historical trends. The electronic system also reduced the ability for attendance record falsification compared to paper sign-in sheets. By modernizing and automating attendance tracking, Uni-Nanny thus addressed many of the major limitations of the traditional paper-based method. The study results clearly demonstrated the Uni-Nanny electronic attendance system's dramatic improvements over paper-based tracking in terms of data quantity, quality, and accessibility. Regarding quantity, Uni-Nanny captured substantially more attendance records on average - 28% more than paper tracking across the modules analyzed. For certain courses, the increase was as high as 90% more records gathered electronically. This major expansion enabled more comprehensive monitoring of attendance patterns and trends.

The research also found significant enhancements to data quality and accessibility with Uni-Nanny. Its instant wireless transfer of records to a central database enabled convenient real-time analysis of attendance data both at the individual class level and across the university. This facilitated easy identification of issues like chronically absent students. The centralized records were a major advantage over paper records, which were decentralized and arduous to collect and consolidate for broad analysis.

The study found 75% of students appreciated the attendance monitoring since it demonstrated the university's care about their academic success. Many desired supportive university intervention for repeated absences. The electronic system also virtually eliminated the common problem of students falsifying signatures on paper sign-in sheets. Staff praised the comprehensive centralized data repository, which eliminated laboriously gathering and consolidating scattered paper records across campus. Automated data capture removed the need for tedious manual entry, freeing up more time for substantive analysis rather than clerical work (Bowen et al., 2005).

In summary, modernizing and automating tracking through Uni-Nanny led to major improvements in attendance data collection, accuracy, accessibility, and analysis compared to paper-based methods. The researchers concluded this type of advanced monitoring approach could greatly strengthen student retention initiatives and data-driven decision-making. Their findings presented a compelling case for institutions to invest in such systems.

The reviewed literature encompasses diverse perspectives on attendance tracking systems in educational settings. The studies examined the transition from traditional paper-based attendance tracking to digital and automated systems, highlighting complexities in understanding the direct impact of digital attendance on academic performance. Factors such as individual attributes, learning environments, and technological influences intertwine to shape student success, challenging the straightforward determination of attendance effects.

Lu and Cutumisu's research underscores the nuanced relationship between attendance and academic performance in modern higher education. Their findings emphasize that while direct attendance may not singularly predict grades, it indirectly influences student performance by fostering self-regulated online learning and formative assessment scores, suggesting technology's potential to supplement reduced physical attendance.

Gustafsson-Wright, Osborne, and Aggarwal underscore the growing significance of real-time data collection in education, particularly in addressing global learning crises and informing data-driven decision-making. They advocate for digital tools' capacity to enhance adaptive management, track progress, and tailor teaching experiences, despite challenges posed by data collection and disparities in access.

Bowen et al.'s study showcases the transformative impact of an electronic attendance system, Uni-Nanny, in higher education. The research outlines the limitations of paper-based tracking methods and highlights Uni-Nanny's substantial enhancements in data quantity, accuracy, and accessibility. This innovative system garnered student appreciation, reduced falsification, and streamlined administrative tasks, advocating for a shift towards advanced monitoring approaches in educational institutions.

Overall, these studies collectively emphasize technology's potential to optimize attendance tracking processes, enhance student engagement, and offer insights for improving data-informed decision-making in educational settings. The findings underscore the importance of investing in innovative digital systems like Uni-Nanny, fostering a compelling case for educational institutions to adopt advanced monitoring approaches to benefit student outcomes and administrative efficiency.

3.2 Machine Learning and Signature Recognition for User Verification

Reliably authenticating user identities is a critical requirement for accurately automating attendance tracking systems. As learning environments continue advancing toward blended and online modalities, verification of attendance becomes increasingly important. While manual validation via paper-based methods is prone to human errors, machine learning approaches analyzed through biometric samples show promise.

One biometric well-suited for educational identity verification is handwritten signatures. Signatures can uniquely identify individuals when captured and analyzed digitally. However, raw signature images contain visually complex patterns that defy simple human interpretation. Distinguishing authentic signatures from expert forgeries poses a significant classification challenge. Fortunately, fields like machine learning and computer vision have made tremendous progress in developing methodologies for analyzing complex biometrics. At the core of signature-based identity verification systems lie interconnected preprocessing, feature extraction, and modeling techniques. Preprocessing methods like masking, normalization and binarization isolate and standardize signatures into comparable inputs. Feature extraction then engineers representations that numerically encode biometric traits, compactly summarizing patterns.

Machine learning models are subsequently trained on vast datasets to learn sophisticated discriminative patterns far beyond human capabilities. Algorithms analyze properly formatted training data to uncover subtle distinctions not intuitively obvious. When optimized through iterative validation, end-to-end systems achieve robust classification abilities.

As educational institutions increasingly investigate digitally-driven solutions, researchers continue exploring technical implementations of attendance tracking enhanced by machine learning-based identity verification. The following sections outline prominent techniques discussed in related literature, with the aim of informing potential technical designs for user authentication supporting online and hybrid course attendance tracking. Detailed explanations establish conceptual foundations for advanced signature analysis techniques and their suitability to education domains.

The paper "*Offline Signature Verification System Using SVM Classifier with Image Pre-processing Steps and SURF Algorithm*" published by *Li Wen Goon* and *Swee Kheng Eng* focuses on developing an effective offline signature verification approach, which faces unique challenges compared to online systems. Specifically, the paper proposes a multi-stage methodology involving image acquisition from a public signature database, image pre-processing techniques such as *binarization*, *Gaussian noise removal* and *thinning*, feature extraction using speeded-up robust features (**SURF**) and bag-of-features modeling, and classification with support vector machines (**SVM**). Two signature verification systems are implemented and evaluated - one without requiring a user ID, and the other verifying signatures individually with entered IDs. The paper also analyzes the classification performance under varying training and testing ratios.

Image Preprocessing

The preprocessing stage forms the foundation of the proposed approach. This phase systematically enhances raw signature images to optimize them for downstream analysis. A multi-step pipeline is designed to purge interfering noise while retaining discriminative structural information. Color images are first binarized, simplifying pixels to clear black and white values. This converts inputs into formats more readily amenable to subsequent refinement. Gaussian filtering is then applied, a specialized low-pass technique that selectively attenuates high-frequency distortions through principled smoothing. Compared to commonly used median filtering, Gaussian filtering is shown to better preserve granular details integral to feature extraction later in the process. Further preprocessing resizes and thins signatures, producing normalized representations of consistent dimensionality and thickness for unified analytical handling. Through strategic application of binarization, noise filtering, resizing, and thinning, the preprocessing regimen purges interfering signals in a controlled manner to yield immaculate, standardized inputs prepared for optimized analysis in later stages.

Feature Extraction

In the feature extraction stage, the paper utilizes *Speeded-Up Robust Features* (**SURF**) detection along with bag-of-features modeling. SURF is employed to efficiently detect localized areas of interest within the preprocessed signature binaries. It quantifies the visual characteristics around densely clustered interest points, annotating this localization information as compact feature descriptors.

These descriptive representations are then aggregated using bag-of-features modeling. This transforms the signature image collection by partitioning it into clusters that can be represented as a "visual word" vocabulary. The signatures are thus conceptualized and discretized into bags of these discriminative visual words.

The extracted bag-of-features vectors undergo classification with Support Vector Machines (SVM). As a supervised learning algorithm, SVM is utilized to learn the intricate decision boundaries between the classes of authentic and forged signatures based on the training dataset. SVM aims to maximally separate these classes through identifying an optimal hyperplane in the feature space.

Two signature verification systems are implemented and evaluated using 10-fold cross validation. One system performs verification without requiring a user ID, while the other classifies individually entered signatures after associating them with a user ID during modeling. Confusion matrices and verification rates are used to assess the classification accuracy.

In summary, this paper presents an effective offline signature verification methodology combining image preprocessing, feature extraction, and machine learning classification. The integrated approach applied strategic preprocessing techniques to optimize signature images before extracting salient features using SURF detection and bag-of-features modeling. These representations were then classified using SVM to discriminate genuine from forged signatures. Experimental evaluation on two verification system configurations demonstrated the approach achieved average accuracies ranging from 71-79%. The results validated that the synergistic multi-stage solution effectively addressed the challenges of offline signature analysis. While offering promising performance, the study additionally suggested future work exploring alternative datasets and models could provide further optimization. Overall, the paper introduced a rigorous technique for biometric identity validation through offline handwritten signature authentication (Goon & Eng, 2021).

In the paper titled "*The Influence of an Electronic Attendance Monitoring System on Undergraduate Academic Success*," Charles Childress explores the impact of implementing an electronic attendance monitoring system on students' academic success. The introduction section of the paper provides the initial context, aims, and hypotheses of the study.

Childress begins by highlighting the importance of attendance in higher education and its correlation with academic success. He acknowledges that traditional manual attendance tracking methods can be time-consuming and prone to errors. In response to these challenges, many institutions have started implementing electronic attendance monitoring systems. These systems use technology, such as biometric scanners or card swipes, to automate the attendance tracking process.

Several research questions and hypotheses are highlighted in the introduction section. These include:

- How does the implementation of an electronic attendance monitoring system affect students' attendance rates?
- What is the relationship between attendance rates and academic performance?
- Does the use of an electronic attendance monitoring system improve students' overall academic success?

Childress mentions that the study is guided by the theoretical framework of social cognitive theory. This theory suggests that individuals' behaviors, such as attendance, are influenced by their beliefs, self-efficacy, and the environment. By applying this framework, Childress aims to understand how the implementation of an electronic attendance monitoring system can influence students' attendance behavior and subsequently impact their academic success.

Also referencing previous studies that have explored the relationship between attendance and academic success, Childress has found a positive correlation between attendance and grades, suggesting that regular attendance is associated with better academic performance. However, there is limited research specifically examining the impact of electronic attendance monitoring systems on academic success.

Childress employs a quasi-experimental design to analyze the influence of the electronic attendance monitoring system on academic success. This design allows for the examination of the relationship between the independent variable (attendance monitoring system) and the dependent variable (academic success) while controlling for potential confounding variables. To gather data for the study, Childress utilizes archival data. This means that the researcher collects pre-existing data from records and databases rather than collecting new data through direct observation or surveys. The use of archival data allows for a large sample size and reduces the potential for biases in data collection.

Childress employs two statistical analyses to examine the relationship between attendance monitoring and academic success: point-biserial correlation analysis and logistic regression analysis.

Point-biserial correlation analysis: This statistical method is used to determine the strength and direction of the relationship between a continuous variable (such as academic success) and a dichotomous variable (such as attendance monitoring). It allows for the calculation of a correlation coefficient that indicates the degree of association between the variables.

Logistic regression analysis: This statistical method is used to examine the relationship between a set of independent variables (such as attendance monitoring) and a binary dependent variable (such as academic success). Logistic regression allows for the estimation of the odds ratio, which indicates the likelihood of the dependent variable occurring based on the independent variables.

These analytical tools provide a rigorous examination of the variables and their significance, allowing the researcher to draw meaningful conclusions about the influence of the electronic attendance monitoring system on undergraduate academic success.

Continuing from the statistical analyses discussed earlier, the influence of an electronic attendance monitoring system on undergraduate academic success, as explored by Charles Childress, becomes evident. Childress's study illustrates the positive impact of such a system, focusing on three crucial aspects: fostering accountability, encouraging regular class attendance, and illuminating the link between increased student engagement, participation, and enhanced academic outcomes.

The role of fostering accountability is significant. In the face of an electronic attendance monitoring system, students become acutely aware that their attendance is not just expected but tracked and recorded. This awareness engenders a sense of responsibility, driving students towards more consistent class attendance. The system's transparency and objectivity eliminate any room for subjective interpretation or discrepancies, thus reinforcing students' understanding of the importance of their academic commitment.

The encouraging regular class attendance is another aspect enhanced by the electronic system. With their attendance monitored and reported in real-time, students are motivated to maintain a consistent presence in class. This real-time feedback aids both students and instructors in identifying attendance patterns and addressing potential issues promptly. Regular class attendance ensures students' active involvement in crucial lectures, discussions, and other academic activities, promoting discipline and a structured learning environment.

Finally, Childress's study underscores the link between increased student engagement, participation, and improved academic outcomes as a result of the electronic attendance monitoring system. Regular attendance provides students more opportunities to engage actively with course material, interact with peers, and partake in class discussions and activities. This active engagement nurtures deeper learning, critical thinking, and knowledge retention. Regular attendance also enables students to establish better relationships with their instructors, seek necessary clarifications, and receive timely feedback. These collective factors significantly contribute to improved academic performance and success.

In essence, the adoption of an electronic attendance monitoring system in undergraduate education positively influences students' academic success. By fostering accountability, encouraging regular class attendance, and boosting student engagement and participation, the system invariably leads to improved academic outcomes. This conclusion underscores the importance of attendance monitoring as a critical tool for enhancing student success, providing crucial insights for educators and administrators in higher education institutions (Childress, 2018).

3.3 Centralized vs Disparate Systems

The article "*Overcoming Data Privacy Issues in Higher Education*" by Meghana Joshi offers an incisive analysis of the data privacy concerns prevalent in higher education institutions. One of the salient points raised pertains to the complications associated with managing disparate systems within these establishments.

Disparate systems, as described in the article, are distinct systems operating within an institution, each with potentially diverse levels of compliance, security protocols, and management needs. These systems often give rise to administrative and security challenges due to the necessity for individual analysis, management, and security measures. This complexity can lead to taxing audits and potential system failures due to the siloed nature of the systems and their varying compliance levels.

Joshi's article underscores the potential exacerbation of data privacy issues in higher education due to the use of disparate systems. It highlights the increased risk of data breaches, the difficulties in adhering to compliance and certification requirements, and the high costs associated with auditing and maintaining multiple systems.

The article also points out that the demand for constant access to administrative services and digital resources can add to the vulnerability of data, emphasizing the need for more unified systems that can better comply with standards and protect privacy.

While Joshi's article does propose *Kivuto* Cloud as a specific solution to these challenges, the broader takeaway lies in the acknowledgment of the challenges associated with managing disparate systems within higher education institutions. Consequently, it suggests the necessity for a more unified approach to system management to enhance data privacy and security.

In continuation with the emphasis on data quality and governance, the article titled "Leveraging Student Information Systems for Data-Driven Decision Making" illuminates the critical role that Student Information Systems (SIS) play within educational settings. Positioned as a foundational tool, SIS acts as a comprehensive database storing multifaceted student-related information, encompassing academic records, demographic details, attendance data, and behavioral patterns.

An emphasis on the pivotal significance of data-driven decision-making within the realm of education permeates the article. Through harnessing data analytics derived from SIS, educational institutions can attain crucial insights into student performance trends, identify at-risk students, and effectively leverage data to enhance teaching methodologies and overall learning outcomes.

Central to the article is the delineation of the instrumental role that SIS assumes in the collection, storage, and analysis of pertinent student-related data. The all-encompassing nature of SIS, encapsulating academic records, demographic data, and behavioral patterns, serves as an invaluable resource for educators and administrators to extract nuanced insights vital for catalyzing improvements in educational outcomes.

Highlighting diverse analytical methodologies applicable to SIS data, the article expounds on descriptive analytics, predictive modeling, data visualization, and machine learning techniques. These tools empower educational institutions to unravel historical trends, predict future trajectories, visually represent data patterns, and uncover latent insights inherent in the dataset.

Moreover, the article underscores the foundational importance of data quality and governance frameworks within the context of SIS. Ensuring the accuracy, security, and consistency of data housed within SIS is not only imperative for deriving credible insights but also for adhering to privacy regulations and preserving the integrity of the dataset.

In essence, the article comprehensively articulates the indispensable role of SIS in facilitating data-driven decision-making processes within educational institutions. By adeptly utilizing the reservoir of information embedded in SIS through robust data analytics, institutions stand poised to enhance their decision-making capabilities, refine pedagogical approaches, and ultimately foster advancements in educational outcomes ("Leveraging Student Information Systems," 2023).

4. Implementation

4.1 System Architecture

The TruMark attendance system is built on a multi-tier architecture using a variety of technologies to ensure robustness, security, and scalability. The system primarily uses HTML, CSS, and JavaScript on the front-end, with Python and PHP on the back-end, and is hosted on an Apache server.

Front-End: The front-end of the TruMark system, developed with HTML, CSS and JavaScript is responsible for providing an intuitive and user-friendly interface. The system's front-end assets are organized within the assets-stellar directory, which contains subdirectories for CSS, JavaScript (JS), SASS, and web fonts. The css directory houses the main styling files, including **main.css** and **fontawesome-all.min.css**, and the js directory contains JavaScript libraries and main scripts that are essential for the dynamic aspects of the website, like user interactions and AJAX requests.

Back-End: The back-end of the TruMark system is built with **Python** and **PHP**. The PHP scripts (courses.php, index.php, login.php, report.php, take_signature.php, update_attendance.php, and update_password.php) are directly located under the main trumark directory and handle various server-side operations, like managing user sessions, handling form submissions, and interacting with the database.

The Python server is responsible for the machine learning aspects of the system. It uses the **scikit-learn** machine learning library to train and test the signature verification model. The trained model is saved as a pickle file (**trained_model_random_forest.pkl**) for future use. The server uses **distance_method.py** and **test_classifier.py** to verify signatures based on the trained model.

Server: The Apache server hosts the TruMark application and handles HTTP requests and responses. The server-side PHP scripts interact with the Apache server to provide dynamic content based on user requests. This multi-tier architecture ensures that each part of the TruMark system can function independently while seamlessly interacting with the other parts. This design allows for efficient troubleshooting and makes the system more maintainable and scalable.

The TruMark system is grounded on an architecture that is designed to handle various data types related to courses, schedules, attendance, and student information. Central to this architecture is a MySQL database named "**Signature**".

The "Signature" database comprises of five primary tables: course_schedule, courses, enrolled, student_attendance, and students. Each table serves a distinct role, managing different aspects of the system's information. However, a brief mention of these tables does not fully capture their importance to the overall system. Each table's structure, purpose, and interaction with the others contribute heavily to the overall functionality of the TruMark system. The following section focuses more on each individual table in more detail, highlighting their role within the "Signature" database.

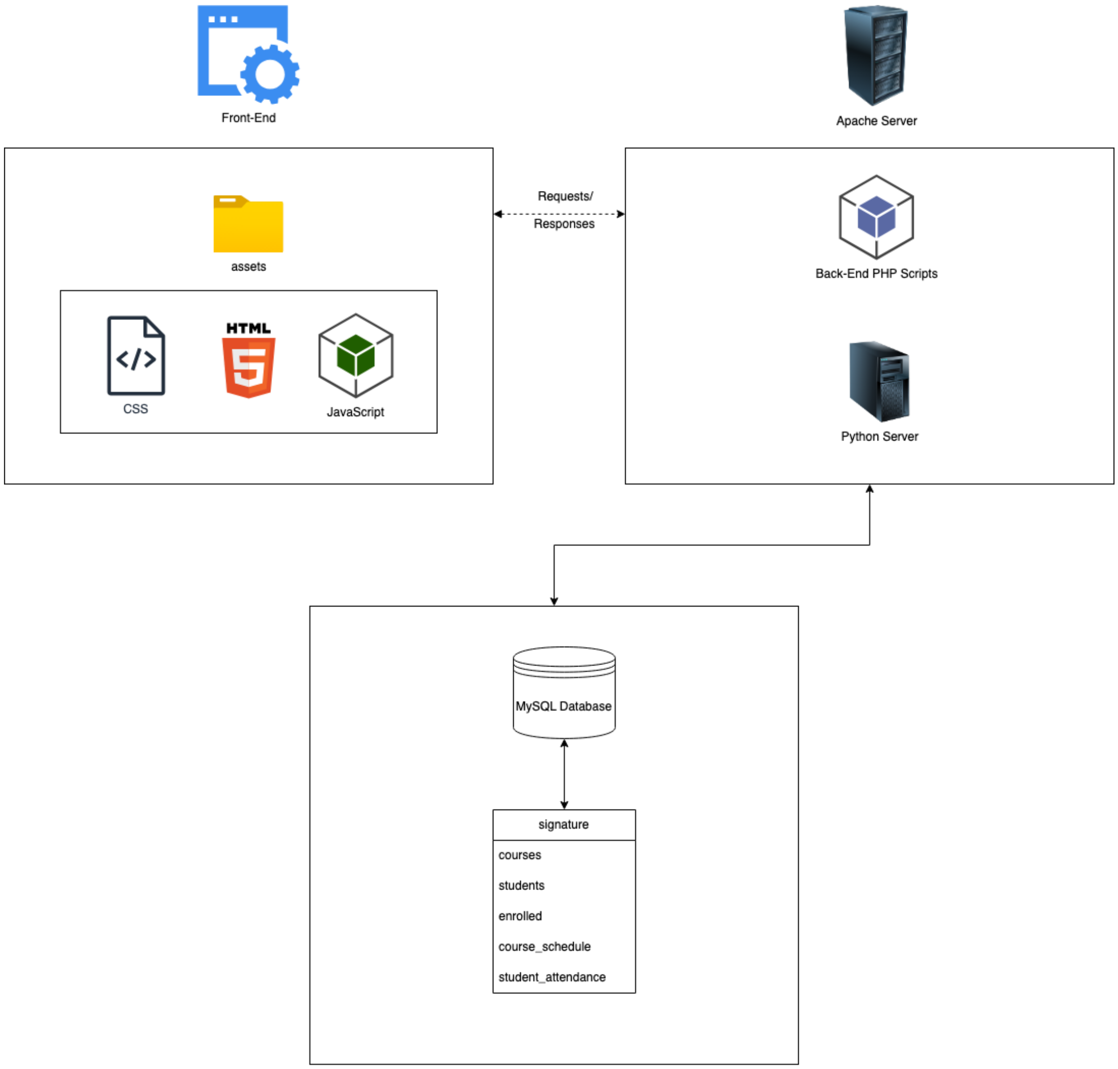

Figure 1.1 System Architecture Diagram, simplified.

4.2 Data Collection

The TruMark system is built around a MySQL database named "Signature". This database is crucial to the operation of the system as it manages and processes various data types related to courses, schedules, attendance, and student information. The database is structured into five primary tables, each serving a distinct purpose:

1. **course_schedule**: This table is a structured set designed to manage the scheduling information for the various courses offered.
   - **schedule_id**: A unique identifier for each schedule. This auto-incremented field serves as the primary key.

- **course_id**: This field creates a clear link between the courses and their schedules.
- **day_of_week**: This enumerated type field holds the day of the week, indicating when a particular course is scheduled.
- **start_time** and **end_time**: These fields of the time data type document the start and end times for the course on the specified day, providing students with a clear understanding of their course timing.

2. **courses**: This table is essentially a catalog containing comprehensive information about the courses on offer.
   - **cid**: A unique identifier for each course, which is auto-incremented and serves as the primary key.
   - **course_name**: This field stores the name of the course.
   - **course_desc**: It provides a detailed description of the course.
   - **course_picture**: The field holds a link or path to an image that represents the course, enhancing the user interface design.
   - **instructor**: Stores the name of the course instructor, providing potential students with an idea of who will be leading the course.

3. **enrolled**: This table tracks student enrollments and encapsulates essential data about student-course associations.
   - **uid** and **cid**: These fields reference the unique identifiers of the students and the courses, respectively, establishing a many-to-many relationship between the students and courses tables.

4. **student_attendance**: This table is instrumental in keeping an accurate record of students' attendance.
   - **attendance_id**: Each attendance record is uniquely identified by this auto-incremented field, which serves as the primary key.
   - **uid** and **cid**: References to the unique identifiers of the student and the course, respectively.
   - **attendance_date**: This field of the date data type records the exact date when the attendance was recorded.
   - **status**: An enumerated type field that helps track the attendance status, with possible values being 'Present', 'Absent', or 'Excused'.

5. **students**: A comprehensive data store that holds a wide range of information about each student.
    - **uid**: A unique identifier for a student record.
    - Additional fields such as username, password, first_name, last_name, email, major, avatar, banner, classification, GPA, level, degree, and advisor hold specific information about a student, providing a detailed overview of the student's profile.

In its current configuration, data for these tables is populated manually to simulate the functioning of the system. However, in a real-world implementation, the data would be collected through the TruMark interface and through integration with university databases. The robust management features of the MySQL database ensure data integrity and security, forming a solid foundation for the TruMark system's operation.

In addition to this relational data, the CEDAR offline signature dataset introduced by Srinivasan et al. (2006) was leveraged to train the machine learning models for signature verification. The CEDAR dataset contains diverse genuine and simulated forged signatures from 55 individuals, supplementing the real student signatures collected.

The variety of samples in CEDAR improves model training and generalization capabilities. The database's robust management ensures data integrity and security, forming a key foundation of TruMark. In its current implementation, data is manually populated but would integrate with university systems in production.

Each of the 55 CEDAR contributors provided 24 genuine samples of their signature, scanned at 300 dpi in grayscale. This resulted in a total of 1,320 genuine CEDAR signatures.

Additionally, some contributors were asked to provide simulated forgeries of other writers' signatures, by attempting to imitate their style. Each contributor produced 8 simulated forgeries for 3 other subjects, resulting in a total of 1,320 simulated forged samples.

4.3 Data Privacy and Security

The current iteration of the TruMark system operates as a prototype, primarily managing manually entered information. This simplified setup, though effective for preliminary demonstrations, underscores the greater complexity we must prepare for when the system is applied in real-world contexts. The importance of meticulous data privacy and security protocols is pivotal, and cannot be overstated — it is the bedrock upon which trust in the system is built.

As we transition the TruMark system towards a model suitable for real-world implementation, the challenges in data handling become increasingly intricate. The system will no longer deal with isolated, manual entries but will instead need to manage a more extensive and diverse dataset. This data, potentially encompassing everything from personal details to academic records, would be of a highly sensitive nature. Thus, secure handling of this information becomes a paramount concern, requiring a comprehensive strategy that ensures both the integrity and confidentiality of the data.

In anticipation of these challenges, a robust approach towards data security and privacy is needed. The implementation of such an approach will involve a series of measures, each designed to address a specific aspect of data security or privacy. These measures won't be static; they will need to evolve in response to the changing data landscape, emerging threats, and updates in regulatory requirements.

In the following sections, we delve into the specific measures that would need to be implemented to ensure the security and privacy of data within the TruMark system. These measures form the cornerstone of the data protection strategy, ensuring that the system remains secure, compliant, and trustworthy as it moves towards real-world implementation.

1. **Encryption**: Sensitive data such as student personal information, course details, and attendance records, would need to be encrypted both at rest and in transit. This means that the data is converted into an unreadable format when stored in the database or transmitted over a network. Only authorized parties with the correct decryption key can access the original data. This measure prevents unauthorized access and protects the data even if the database or the network is compromised.
2. **User Authentication**: Strong user authentication measures would need to be put in place. These could include enforcing strong password policies, implementing multi-

factor authentication, and using secure session management. Such measures would ensure that only authorized individuals can access the system and the data it contains.

3. **Access Control**: Implementing role-based access control (RBAC) could help further ensure that users only have access to data relevant to their role. For instance, a student user would not be able to access other students' attendance records, or an instructor user would only have access to the course data relevant to them.
4. **Auditing and Monitoring**: Regular auditing and monitoring of the system would be crucial to identify any unauthorized access or unusual activity. Automated alerts could be set up to notify system administrators of any potential security breaches.
5. **Data Protection Compliance**: The system would need to be compliant with data protection laws and regulations such as the General Data Protection Regulation (GDPR) in the EU or the California Consumer Privacy Act (CCPA) in the U.S. This would involve implementing measures such as data minimization, purpose limitation, and ensuring the rights of the data subjects.
6. **Privacy Policy Developmen**t: A comprehensive privacy policy would be developed. This policy would inform users about how their data is collected, stored, used, and protected. It would also explain their rights as data subjects, such as the right to access their data, the right to correct inaccuracies, and the right to have their data deleted.

These measures would help ensure that the TruMark system handles data securely and respects the privacy of all users. As data privacy and security are dynamic and constantly evolving fields, the system would also need regular security updates and privacy assessments to stay ahead of potential threats.

4.4 Machine Learning Algorithms

TruMark employs a combination of advanced image processing techniques and machine learning algorithms to achieve accurate and reliable predictions. This section details the chosen algorithms, their functions, and how they contribute to TruMark's overall operation.

Image Processing and Feature Extraction

The initial phase in the application of machine learning algorithms is the processing and transformation of raw input data. In the context of TruMark, the input data consists of images which must be adapted into a format that can be effectively interpreted and utilized by the machine learning algorithm.

The Python Imaging Library (PIL) is employed to manipulate image data. The first operation implemented using PIL is the conversion of images to grayscale. This transformation helps reduce the data's complexity by moving from a space that potentially encompasses millions of colors to a simpler one, consisting solely of grayscale values. Despite its simplicity, grayscale conversion retains the critical characteristics of the images necessary for subsequent processing steps.

Subsequent to grayscale conversion, the images are subjected to 'trimming', a process that entails cropping a uniform amount of pixels from all edges. This operation is instrumental in focusing the algorithm's attention on the central parts of the images where the majority of the significant features are likely to reside.

Following the trimming process, the images are transformed into two-dimensional arrays. Each cell within these arrays corresponds to an individual pixel in the image and contains a binary value: '0' represents white pixels (interpreted as the background), while '1' signifies black pixels (interpreted as the foreground). This transformation facilitates the machine learning algorithm's ability to process and analyze the images by translating the visual data into a numerical format.

Noise reduction is a critical step in data preprocessing. Therefore, a filtering process is applied to eliminate rows and columns in the arrays that contain fewer foreground pixels than a pre-defined threshold. This process aids in reducing noise and enhances the focus on the salient components of the image.

The final step in preprocessing is resizing the arrays to ensure a uniform size. This standardization guarantees that all images, irrespective of their initial sizes, are represented consistently. This uniformity is key for the machine learning algorithm since it allows for a standardized interpretation of images, enabling accurate predictions based on the processed image data.

In summary, the image processing and feature extraction phase is an extensive, multi-stage process involving the transformation of raw image data into a format suitable for machine learning operations. Each step, from grayscale conversion to resizing, aims to simplify and standardize the input data, thereby setting the stage for the optimal operation of the machine learning algorithm and the generation of reliable results.

Feature Engineering

The stage subsequent to image processing is feature engineering, a critical phase in the machine learning pipeline. This process involves the extraction of quantifiable attributes or characteristics from the processed images. These attributes, known as features, are integral to the functionality of machine learning models. The accuracy and reliability of these models are significantly dependent on the selection and formulation of appropriate features.

In the context of TruMark, the primary feature used is a weighted sum of the pixel values for each row in an image. Each pixel's value is multiplied by its corresponding column index, which acts as a weight, and the products are summed across each row. This approach incorporates spatial information into the features, effectively considering both the presence of a pixel ('ink') and its relative position within the row.

The result of this computation is a single numerical value for each row, representative of the distribution of 'ink' within that row. This operation is repeated for every row in the image, yielding a feature vector that encapsulates the horizontal 'ink' distribution for the entire image. This feature vector, which indirectly captures the structural characteristics of the characters in the image, forms the input for the subsequent machine learning model.

The strategy of focusing on the 'ink' distribution for feature engineering proves to be particularly effective in image classification tasks. It capitalizes on the inherent variability between different characters, as the 'ink' distribution is likely to be distinct for different characters. This method, therefore, provides a robust foundation for the machine learning algorithm to differentiate between various character classes.

During feature engineering, raw data undergoes a transformation into a set of features that succinctly capture the essential characteristics of the data. This transformation significantly enhances the efficiency and accuracy of the machine learning models. For TruMark, the computation of a weighted sum of pixel values for each row in an image forms the basis of this transformation, resulting in a set of insightful features for the subsequent stages of the machine learning pipeline.

- **Multilayer Perceptron (MLP) Classifier**

With our features prepared, we can now feed them into our machine learning algorithm, the Multilayer Perceptron (MLP) Classifier. The MLP Classifier is a type of artificial neural network -- a machine learning model inspired by the human brain.

The MLP Classifier consists of layers of nodes, or 'neurons', with each layer taking input from the previous one and providing output to the next. The first layer is the input layer, where we feed in our features. The last layer is the output layer, which provides the final prediction. Between them are one or more 'hidden layers' where the actual computation takes place.

These computations involve taking weighted combinations of the inputs and applying a non-linear function known as the Rectified Linear Unit (ReLU) to the result. The weights are initially set at random, and the purpose of training the network is to adjust these weights based on how well the network is predicting the correct answers.

Training involves a process called backpropagation, where the network's errors are propagated backward from the output layer to the input layer, adjusting the weights along the way. We use the 'adam' solver for this process, which is an efficient variant of stochastic gradient descent -- a method for finding the minimum of a function.

The MLP Classifier contributes significantly to TruMark's ability to analyze and interpret complex datasets. By learning from the patterns in the training data, it enables TruMark to accurately classify new instances and provide meaningful insights.

Saving and Loading the Model

In the context of Machine Learning (ML), a model is trained to learn from data and make predictions or decisions without being explicitly programmed to do so. The process of training a model can be computationally intensive and time-consuming, especially with large datasets. Once a model is trained, it's often necessary to save the model so that it can be reused later without having to retrain it, saving time and computational resources.

In the Python scripts provided, a machine learning model is trained, saved to disk, and later loaded from disk. This is done using the sklearn.neural_network.MLPClassifier class from the sklearn library and the joblib library.

- **Saving the Model**

The saving of the model is done in the **train_classifier.py** script. The script begins with importing necessary libraries and loading a dataset from a CSV file. The dataset is assumed to be in a format where the last column is the target variable (the values to be predicted), and all the preceding columns are the features (the input data used to predict the target).

```
import pandas as pd

from sklearn.neural_network import MLPClassifier

import joblib

import sys


# Load dataset

data = pd.read_csv(sys.argv[1])


# Assuming the last column is the target variable and the rest are features

X = data.iloc[:, :-1].values

y = data.iloc[:, -1].values
```

The script proceeds to create an instance of the MLPClassifier class, which represents a Multi-Layer Perceptron (a type of Artificial Neural Network) classifier. This classifier will be trained on the dataset.

```
clf = MLPClassifier(hidden_layer_sizes=(100,), activation='relu', solver='adam', alpha=0.0001,

                    batch_size='auto', learning_rate='constant', learning_rate_init=0.001,

                    max_iter=200, shuffle=True, random_state=None)
```

After training the classifier on the dataset using the fit method, the model is ready to be saved. The joblib.dump function is used to serialize the trained model to a file. The name of the file is '**trained_model.pkl**'.

```
# Train the model on the entire dataset

clf.fit(X, y)



# Save the trained model with the '.pkl' extension

joblib.dump(clf, 'trained_model.pkl')
```

The model is now stored on disk and can be used later without needing to retrain it.

- **Loading the Model**

The loading of the model is done in the test_classifier.py script. The script begins with importing necessary libraries, including the joblib library used to load the model from disk.

```
import joblib

import pandas as pd

import distance_method

import sys

import warnings
```

```
import os

from sklearn.exceptions import InconsistentVersionWarning
```

The script defines a function, ***classify***, which loads the model from disk and uses it to make a prediction for a provided instance of data (a line from a CSV file).

The loading of the model is done using the joblib.load function. The loaded model is then used to predict the class of the provided data instance.

```
DIRECTORY = os.path.expanduser('~/server/')

FILE_NAME = os.path.join(DIRECTORY, 'trained_model_random_forest.pkl')



# Load the saved model

loaded_clf = joblib.load(FILE_NAME)
```

Once the model is loaded, it is used to make a prediction for the provided data instance.

```
# Predict the label for that instance

prediction = loaded_clf.predict(data_instance)
```

The prediction result is then returned as a string. The model can now be used to make predictions on new data without having to be retrained. This process significantly speeds up the prediction process, especially for large models and datasets.

In summary, the Python scripts provided demonstrate the process of saving a trained machine learning model to disk, and loading that model from disk to make predictions. This process is crucial in many machine learning workflows, as it allows models to be reused without needing to be retrained, saving both time and computational resources.

- **Server Script**

Finally, we have a server script that uses the Python's built-in HTTP server library to receive images, process them using the methods mentioned above, classify them using the pre-trained model, and return the predictions.

In summary, TruMark utilizes a combination of image processing techniques and the MLP Classifier machine learning algorithm to deliver a robust and reliable service. From preprocessing the raw images to making final predictions, each step is crucial in turning raw data into valuable insights.

5. Results and Recommendations

In designing the TruMark system, two pre-processing techniques were considered and tested - masking and distance methods - for evaluating different machine learning algorithms to determine the optimal approach for user verification through signature analysis.

The accuracy results for the evaluated algorithms using each pre-processing method are shown below:

Algorithm	Masking Method	Distance Method
Decision Tree	60%	49%
Random Forest	73%	80.75%
SVM (Support Vector Machine)	45%	69%
Gaussian Naive Bayes	42%	75%
KNN (K-Nearest Neighbor)	71%	59%
Neural Networks	6%	22%

Table 1: Algorithm Performance

Note: Higher percentages indicate better performance.

Based on having the highest accuracy, the Random Forest algorithm with distance pre-processing was selected as the optimal approach for user verification in the TruMark system. The accuracy of **80.75%** surpassed other combinations of algorithms and pre-processing techniques.

For Truman State University, TruMark represents an important step toward modernizing daily attendance tracking by applying innovative machine learning and signature analysis methods. It provides a more automated, secure, and convenient system over traditional manual processes.

One recommendation is exploring integration of TruMark with the University's existing TruView platform. This could allow leveraging TruMark's user verification capabilities while avoiding adoption of disparate systems. The results of this project validate the value of the approach and its potential for continued enhancement within the Truman ecosystem.

While the TruMark prototype demonstrates promising capabilities for automated attendance tracking through machine learning and signature analysis, the current implementation faces some limitations that could be addressed in future work.

One significant limitation is the small dataset size used for training and evaluating the machine learning models for signature verification. The prototype was developed using a dataset of only 24 signature samples per user. Expanding the training data to include more users with more diverse signatures would help improve the generalization performance of the models and accuracy on new signatures. Ideal future work could collect thousands of signature samples from hundreds of actual students and faculty at Truman to better represent the real-world use case.

Additionally, user testing of TruMark has been limited in scope. Testing TruMark in an operational pilot program with campus classes would provide valuable insights into real-world performance and usability. The constraints of the academic calendar and schedule introduce logistical challenges for effectively launching such a pilot evaluation. But collecting authentic usage data and user feedback from real students logging attendance will be an important step before potentially scaling TruMark campus-wide.

There are also abundant possibilities for enhancements and new features building on top of the core TruMark platform. Developing a mobile app version could greatly expand accessibility and convenience of attendance tracking for students and faculty. Rather than needing to log in on a web browser, having TruMark available via a smartphone application would enable quick attendance check-in regardless of location. This mobile approach could leverage built-in features like GPS location tagging or facial/biometric authentication available on modern devices.

Integrating TruMark with university course scheduling systems or the BrightSpace learning management system could enable useful automated workflows. Rosters could automatically populate attendance sheets and tracking could interface with assignment records. Notifications and messaging could be added to inform students of attendance requirements or alert faculty of absences. Providing analytical dashboards or visualizations could give instructors deeper insights into class attendance trends.

Long-term, the identity verification techniques developed for TruMark may have wider applications beyond attendance tracking. The signature analysis and machine learning models could potentially be adapted to support student login for campus facilities, transactions and payments, access to records, and other scenarios requiring trusted authentication. Exploring these extended use cases could uncover additional value for Truman beyond modernizing day-to-day attendance management.

The current iteration of the TruMark system, while serving its purpose in the prototype phase, faces certain limitations that need to be acknowledged and addressed. The most prominent among these is the reliance on manual processes for data collection and management. This manual approach, while allowing for precise control in the system's early stages, presents several disadvantages that could hinder its performance in larger, real-world deployments.

Firstly, manual data entry inherently carries the risk of human error, which could lead to inaccuracies in the data. Secondly, the manual process is time-consuming and labor-intensive, which could pose scalability issues as the system grows and needs to handle larger volumes of data. Lastly, the current system lacks real-time responsiveness; changes in data, such as updates in student records or course details, are not immediately reflected in the system. This lag potentially compromises the system's accuracy and efficiency.

To address these limitations and prepare the TruMark system for large-scale deployment, we can envision several future improvements. Most notable is the automation of the data collection and management process. This could involve integrating the TruMark system with existing university databases, enabling the automatic importation and regular updating of student and course data. This integration would foster real-time synchronization between the system and the databases, minimizing potential data discrepancies, and eliminating the need for manual data updates.

However, it's important to note that these improvements are not the final destination but rather milestones in the TruMark system's ongoing journey. As the system is exposed to real-world scenarios, it will inevitably face new challenges and requirements. Future enhancements will therefore need to be responsive to these evolving needs, as well as to user feedback, technological advancements, and updates in data protection legislation and educational policies.

In essence, the evolution of the TruMark system is a continuous process marked by constant learning, adaptation, and iterative improvement. The aim is delivering an optimized, secure, and user-friendly solution that best serves its users. The TruMark prototype serves as an important foundational demonstration of automated attendance tracking, but there are many promising directions for refinement and expansion that align with this ongoing evolution. Improving the machine learning approach through more robust datasets and real-world testing is critical to this process. Building out the feature set and integration with Truman's existing technical infrastructure can increase the value proposition over time. And supplementing the attendance use case with new applications of user verification could demonstrate the versatility of the underlying technology as TruMark develops. Using TruMark as the starting point, there are many potential avenues to improve and evolve automated systems to enhance academics and administration at Truman State University through this continuous journey of progress.

6. Conclusions

6.1 Importance in Educational Technology

TruMark demonstrates the immense potential for leveraging innovative technologies like machine learning and signature analysis to dramatically enhance academic administration processes. By automating the previously manual and tedious task of taking daily attendance, TruMark provides tremendous convenience and efficiency gains while still ensuring data integrity through its underlying security features.

This aligns with broader trends and a growing recognition across higher education of the power of utilizing advanced technologies to enrich and optimize the educational experience for both students and faculty. Automated attendance tracking is just one example - there are abundant opportunities to leverage technology to streamline and upgrade antiquated manual workflows across university administration.

Specifically, this project highlights the value of developing customized systems tailored to a university's unique needs, rather than taking a one-size-fits-all approach. TruMark was built from the ground up with Truman State's specific requirements, use cases, constraints, and stakeholders in mind. This exemplifies why educational institutions can derive great benefit from investing in specialized tools optimized for their particular environment, resources, and student population.

In addition, TruMark provides an excellent example of how technology, when thoughtfully designed, can be made intuitive and user-friendly even for complex tasks like attendance tracking across courses. Students and faculty alike can seamlessly interact with TruMark's system without extensive training or friction. TruMark demonstrates how properly designed interfaces and experiences can make technology feel empowering rather than imposing for end users.

Looking ahead, the aim of this project is to pave the way for Truman State to continue exploring innovations in automating other aspects of academics and administration. The possibilities are abundant, whether in streamlining scheduling, advising, campus accessibility, notification systems or more. TruMark represents an important first step on an ongoing journey of leveraging secure, tailored educational technologies to create a more intuitive, empowering environment fostering student success.

6.2 Relevant Coursework

Throughout my computer science coursework, I gained critical knowledge and skills that I directly applied when developing the TruMark system for this capstone project.

- **CS180 - Foundations of Computer Science I** provided my first introduction to programming and thinking algorithmically to solve problems. I learned basics like data types, variables, control structures, loops, functions, and arrays. This allowed me to write the core logic and workflows for components like TruMark's user authentication, course registration, attendance taking, and administrator dashboards. The basic programming fundamentals were essential building blocks.

- In **CS181 - Foundations of Computer Science II**, I leveled up my skills with object-oriented programming in an event-driven context. Learning about object classes, inheritance, and event handlers enabled me to architect TruMark's graphical user interface and interactive web experience. For example, I defined custom object classes for the user profiles, courses, and attendance records. The OOP approach helped organize my code and simplify cross-component interactions.

- **CS260 - Object-Oriented Programming and Design** deepened my OOP skills through concepts like abstraction, encapsulation, and polymorphism. I learned design patterns like MVC and Observer that I applied to decouple TruMark's backend logic from the frontend UI. This enabled effective modular and extensible code. I also used inheritance and interfaces extensively for reusing and extending class behaviors. The OOP focus pushed me to think about flexible software architecture.

- Through **CS310 - Data Structures and Algorithms**, I learned how to leverage data structures like arrays, linked lists, stacks, queues, trees, and graphs to efficiently store and access data. This allowed me to optimize TruMark's algorithms by managing complexity.

- **CS315 - Internet Programming** provided web development skills in HTML, CSS, JavaScript, PHP, and more that were crucial for implementing TruMark's frontend as a dynamic web application. I learned web protocols like HTTP alongside AJAX techniques for smooth page transitions. My frontend design skills grew tremendously.

- **CS370 - Software Engineering** gave methodologies for properly documenting requirements, diagramming UML models, writing test cases, and modular code organization. I followed these best practices to systematically design, implement, and test TruMark's multi-component system. This guidance was invaluable for coordinating and implementing this project.

- **CS430 - Database Systems** allowed me to model TruMark's backend data storage needs with ER diagrams and database schema design, and gave me the foundation of MySQL database knowledge.

- **CS470 - Computer Networks** provided networking foundations I used when architecting TruMark's client-server web architecture. I incorporated concepts like TCP/IP, DNS, and web security protocols when making architectural decisions for reliable uptime. This guided key design choices.

By extensively leveraging all my computer science coursework, I gained the full stack of expertise to successfully tackle TruMark's development. The specific skills in programming, algorithms, databases, software design, and networks came together to enable creating this complex system.

7. References

Bowen, E., Price, T., Lloyd, S., & Thomas, S. (2005). *Improving the quantity and quality of attendance data to enhance student retention*. *Journal of Further and Higher Education*, *29*(4), 375–385. https://doi.org/10.1080/03098770500353714

Childress, C. (2018). *The Influence of an Electronic Attendance Monitoring System on Undergraduate Academic Success* (Doctoral dissertation, University of Southern Mississippi). Aquila - University of Southern Mississippi Digital Commons. https://aquila.usm.edu/dissertations/1560/

Goon, L. W., & Eng, S. K. (2021). *Offline Signature Verification System Using SVM Classifier with Image Pre-processing Steps and SURF Algorithm. Journal of Physics: Conference Series, 2107,* 012069. https://doi.org/10.1088/1742-6596/2107/1/012069

Gustafsson-Wright, E., Osborne, S., & Aggarwal, M. (2022, November). Digital tools for real-time data collection in education. Brookings. https://www.brookings.edu/articles/digital-tools-for-real-time-data-collection-in-education

Joshi, M. (2019). *Overcoming data privacy issues in higher education. Kivuto Blog.* https://kivuto.com/blog/overcoming-data-privacy-issues-in-higher-education/

Kim, E. (2020, October 17). How time is being spent in class. *The Bull's Eye*. https://dbbullseye.com/2020/20995/

Leveraging Student Information Systems for Data-Driven Decision Making. (2023). Data Analytics.https://www.classter.com/blog/data-management/data-analytics/sis-for-data-driven-decision-making/

Lu, C., & Cutumisu, M. (2022). Online engagement and performance on formative assessments mediate the relationship between attendance and course performance. *International Journal of Educational Technology in Higher Education, 19*(1). https://doi.org/10.1186/s41239-021-00307-5

Srinivasan, H., Srihari, S. N., & Beal, M. J. (2006, December 15). Machine learning for signature verification. CEDAR Signature [Dataset]. Retrieved from https://paperswithcode.com/dataset/cedar-signature