SMARTSIZE AI – MACHINE LEARNING BASED CLOTHING SIZE RECOMMENDER

A Project Report
Submitted in partial fulfillment of the
requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

DEPARTMENT OF COMPUTER SCIENCE and Engineering

By

HARSHINI S. R- 2420030113

GIRIDHAR – 2420030171

SATYADEV – 2420090120

Under the Esteemed Guidance of

LECTURER NAME
Dr. Siripuri Divya

(DEEMED TO BE UNIVERSITY)

Koneru Lakshmaiah Education Foundation

(Deemed to be University estd. u/s. 3 of the UGC Act, 1956)
Off-Campus: Bachupally-Gandimaisamma Road, Bowrampet, Hyderabad, Telangana - 500 043.
Phone No: 7815926816, www.klh.edu.in

K L (Deemed to be) University

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Declaration

The Project Report entitled “SMARTSIZE AI” is a record of bona fide work of **team members, 2420030113-Harshini S. R, 2420030171 – Giridhar, 2420090120 - Satyadev,** submitted in partial fulfillment for the award of B. Tech in Computer Engineering to the K L University. The results embodied in this report have not been copied from any other department/University,/Institute.

2420030113 - Harshini

2420030171 - Giridhar

2420090120 - Satyadev

K L (Deemed to be) University

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Certificate

This is certified that the project-based report entitled "SMARTSIZE AI" is a bonafide work done and submitted by **Harshini (2420030113), Giridhar (2420030171), Satyadev (2420090120)** in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in Department of Computer Science Engineering, K L (Deemed to be University), during the academic year **2024-2025.**

Signature of the Supervisor

Signature of the HOD

Signature of the External Examiner

ACKNOWLEDGEMENT

The success in this project would not have been possible without the timely help and guidance rendered by many people. We wish to express our sincere thanks to all those who have assisted us in one way or the other for the completion of my project.

Our greatest appreciation to our guide, Dr. Divya Siripuri, Department of Computer Science, who cannot be expressed in words for her tremendous support, encouragement and guidance for this project.

We express our gratitude to **Dr. B. Venkateshwara Rao**, Head of the Department for Computer Science Engineering, for providing us with adequate facilities, ways, and means by which we are able to complete this project-based Lab.

We thank all the members of teaching and non-teaching staff, and also those who have assisted me directly or indirectly, for the successful completion of this project. Finally, I sincerely thank my parents, friends, and classmates for their kind help and cooperation during my work.

Harshini - 2420030113

Giridhar - 2420030171

Satyadev - 2420090120

TABLE OF CONTENTS

S.No	Contents	Page no
1	Abstract	
2	Introduction	
3	Literature survey	
4	Client meetings	
5	Hardware and Software requirements	
6	Implementation	
7	Experimentation and Code	
8	Results	
9	Conclusion	
10	References	

ABSTRACT

1. Project Overview

This paper presents Smart Size AI as an artificial intelligence-powered sizing platform of the next generation, aiming to solve one of the major problems in the global clothing industry: accurately predicting and recommending the right size across various consumer populations. The system applies machine learning; it combines computer vision and large-scale data analytics to enhance people's fit within online retailing stores or offline ones. The project originates from interdisciplinary research on fashion technology, AI, UX design, data privacy, as well as supply chain optimization.

The world in which almost all aspects of life are nearly dominated by e-commerce and offers a wide market to the international audience makes it imperative for you to look for a remedy that will finally solve the issue of size normalization, cultural fit variations, and the constant evolution of textiles and wearables. Smart Size AI is the fusion of systems-approach design, software engineering development methods, and iterative stakeholder management techniques into building a recommendation engine that is very accurate, scalable, modular, easily upgradable for new markets, and evolving uses system

2. Problem Statement

E-commerce innovations in practice are too often undermined by the inability to recommend proper fits, resulting in high product return rates, low customer satisfaction, unsold inventory, waste, and negative environmental impacts. Classic sizing systems-either static size charts, rule-based algorithms, or customer guesswork-are inadequate in their capacity to accommodate inter-individual variability because of genetics, lifestyle, and evolving fashion preferences.

Poor-fit returns make up approximately 30% of all online apparel purchases globally, causing tremendous losses for retailers and manufacturers, while promoting consumer frustration. This situation worsens in marketplaces catering to multicultural populations with varying anthropometric distributions.

3. Project Motivation and Goals

Smart Size AI is inspired by the vision to reduce the number of misfit returns while enhancing the confidence of online shoppers and making it easier for retailers in inventory management. The objective of this project is three-fold:

- To engineer a sizing platform that would minimize return rates due to improper fits.
- To make e-commerce apparel more inclusive by catering to users across a wide spectrum in terms of body shape, size, and accessibility needs.
- To enable data-driven design and production for apparel manufacturers and brands.
- To ensure user privacy by applying best practices in ethical AI and regulatory compliance.
- To build the flexible base for future expansion with augmented-reality fitting and biometric adaptive clothing.

4. Methodological Foundation

Smart Size AI uses a microservices structure to divide tasks between data intake, modeling, suggestion sending, and analytics display. The backend uses microservice APIs (made with Node.js), an AI inference engine (Python, TensorFlow, scikit-learn), and data storage (MongoDB, Redis). Real-time and batch data streams are synced using task schedulers and data pipelines.

The suggestion system works by combining three sources: user measurements, past purchases, and image-based body extraction using deep learning. Data merging, ensemble learning, dimensionality reduction, and feedback are used to improve prediction accuracy and adjust to changes in user actions and clothing stock.

5. Background

Smart Size AI's design is based on current research. This includes work on convolutional neural networks for analyzing 3D body scans, methods of collaborative filtering like those in TrueFit and Fit Analytics, and machine learning setups that keep data private. To handle problems with limited data and changing data patterns, the platform uses data changes, transfer learning, and specific learning rates. Studies point out that different body types must be included, and data leaks are a risk. These issues are key concerns in our design. Studies emphasize the importance of inclusivity for underrepresented body types and the risks posed by data privacy breaches; both are central concerns addressed in our architecture.

6. System Features & Innovations

Smart Size AI differentiates itself through the following innovations:

- **Demographic Coverage:** The ability to handle multiple age groups, genders, and ethnic backgrounds, using customizable size charts and flexible data representation.

- **Body Shape Analytics:** Utilization of advanced keypoint detection and pose estimation algorithms for high-precision dimension extraction from user images.

- **Privacy-First Design:** End-to-end encryption, strict access controls, and user data minimization are default principles.

- **Retailer Tools:** Granular analytics for sizing trends, inventory optimization, and return risk prediction.

- **Multimodal Input:** Support for manual measurements, image uploads, and soon, smartphone-based AR scanning.

7. Technical Implementation

The backend microservices work in conjunction with a real-time message broker to enable scalable concurrent processing of recommendations. Model serving is containerized with Docker for rapid deployment across test, staging, and production environments. Data preprocessing involves imputation, noise filtering, normalization, and one-hot encoding for categorical variables. Ensemble approaches, random forests, XGBoost, and neural collaborative filtering,form the predictive backbone, while model evaluation leverages k-fold cross-validation.

On the frontend, a React-based user interface supports multi-modal onboarding, adaptive accessibility components, and dynamic visualization of fit recommendations. Integrations with third-party e-commerce APIs enable seamless deployment in retailer contexts.

8. Dataset Design and Augmentation

Smart Size AI draws on multiple dataset sources: synthetic anthropometric data, anonymized retailer databases, public size charts, and real user inputs. Extensive use of data augmentation, rotation, scaling, and synthetic user generation,alleviates overfitting and enhances robustness, especially in privacy-constrained environments.

9. Stakeholder Engagement

The development lifecycle includes continuous feedback from key stakeholders: users (via surveys, feedback forms, AB testing panels), retailers (feature prioritization, analytics needs), and regulatory experts (compliance audits, privacy reviews). Requirement traceability matrices, sprint backlogs, and regular client meetings ensure alignment with commercial and ethical objectives.

10. Challenges and Risk Mitigation

Throughout the development and deployment of Smart Size AI, several primary risks are systematically addressed:

- **Data Privacy:** Compliance with GDPR, CCPA, and local regulations.

- **Bias Mitigation:** Algorithmic audits for demographic fairness and bias correction.

- **Scalability:** Use of elastic cloud infrastructure, automated resource scaling.

- **Robustness:** Multi-layered validation and exception handling for atypical and adversarial input cases.

11. Experimental Validation

A series of validation experiments assesses model performance across major apparel categories (tops, bottoms, dresses, outerwear). Metrics include prediction accuracy, model confidence calibration, reduction in return rates, and user satisfaction rates. Experimental design includes control groups (traditional sizing) and treatment groups (Smart Size AI), with pre-/post-analysis.

12. Impact Analysis

Smart Size AI demonstrates significant improvements in major KPIs:

- Average sizing accuracy increases by 20-30% over standard charts.
- Retailer return rates decrease by 15–35% depending on implementation scope.
- User satisfaction in fit and purchase confidence sees a measurable lift in NPS and repeat engagement.

13. Future Directions

The plan for Smart Size AI includes extending to AR-enabled fitting rooms, real-time fitting visualization, global clothing database profiles, and open APIs to integrate with future virtual retail technologies. Research plans also include zero-knowledge body measurement protocols and research in federated learning to enhance privacy even further.

14. Summary of Contributions

Smart Size AI reflects a synthesis of theory and applied innovation and provides the following principal contributions:

- An AI sizing architecture that is scalable and that uses a modular approach.
- A robust design for data-pipelining, including privacy features.
- Literature review and industry benchmarking with significant coverage and depth.
- A full-stack implementation, including cloud-native functions.
- Open-source modules for extension by community or academic groups.

15. Conclusion

This project introduces a fully integrated strategy to address the issue of sizing apparel and develops solutions based on contemporary artificial intelligence aspects. The system captures real transformational value for consumers, retailers, and the larger sustainability objectives of the industry while showing academic value and commercial viability.

INTRODUCTION

1. Introduction: A Global Apparel Sizing Problem

Apparel sizing presents a difficult problem on both a technical and social level. Current sizing systems are rooted in old surveys and limited standards. As the clothing market becomes increasingly global and online shopping expands, people struggle to find clothes that fit. This results in many returns and hurts confidence in buying clothes online. The clothing business deals with increased costs, ecological problems caused by waste, and a pressing demand for custom and eco-friendly answers.

2. A Solution Overview

Smart Size AI aims to change how clothes sizes are matched to shoppers. It uses computer vision, deep learning, and analytics to process images and measurements for accurate suggestions. Its flexible design allows it to be added to different retail settings, like online stores, smart mirrors, and mobile apps. The system is built for quick updates, testing, and customized features for different customers.
Its core features include:

- Precise measurement: Algorithms pull exact measurements from user images or device data.
- Machine Learning: Models get better over time with new info, improving suggestions for everyone.
- Data Use: Clothes details are matched with sales data and reviews to improve the recommendation system.
- Privacy: It follows global rules and gives users control through open tools.
- Smart Size AI works for online shoppers, in-store kiosks, mobile AR fitting, and retailer API access. The goal is to build trust between brands and shoppers, while making fashion accessible, sustainable, and more personal.

3. Interdisciplinary Functions

Smart Size AI blends fashion tech, data study, supply chain study, and how people and computers work together:

- Fashion know-how informs customization, material behavior, and algorithm changes for new styles and materials.

- Data study, AI, and stats support prediction and fairness, with constant checks and bias fixes.
- How people and computers work together improves ease of use through varied onboarding, feedback, accessible designs, and interfaces based on behavioral science.
- Rules and ethical guidelines (like GDPR and CCPA) support all system functions, focusing on permission and limited data use.

4. Problem Definition and Research Questions

Fashion sizing goes beyond just numbers. It's about turning body and clothing info into suggestions that match what people like and feel good wearing. Key questions include:

- How can tech help match personal taste with good predictions, and also be clear about why certain things are suggested?
- What types of information, such as pictures, sizes, or past buys, are most useful, while still protecting people's privacy?
- How can we be sure that the suggestions are fair, open, and without any unfair bias?

Some hard tech issues are having too little data, wrong data, unusual data, data that is not balanced, and learning from what people say over time. Ethics are very important, not just in making suggestions, but also in how data is handled and how much control people have.

5. Market and User Research

Global apparel markets are different due to region-specific preferences. Groups like young people, older adults, athletes, and those needing particular sizes need specialized sizing, which pushes brands to update old systems. User profiles in Smart Size AI shape the interface and algorithms, focusing on ease of use, reliability, quick responses to feedback, and how relevant it is to the user. Lifestyle and where people live also inform how users are divided and what's suggested to them.

6. Standard Sizing Systems

Worldwide sizing standards give a framework but aren't always followed, causing problems for brands and confusion for shoppers, especially across countries. Smart Size AI changes to fit both general size charts and real sales info. They focus on making things specific to each person using feedback, and easy-to-change software methods let more people come up with sizing ideas instead of just those with costly equipment.

7. Body Data and How It's Taken

To get sizing right, you need good body data taken either directly (with tape measures) or indirectly (photos, scans). Smart Size AI uses a mix of methods to make sure data is correct, lets users fix errors, and tries to include everyone. They add to data and use fake profiles to be fair, and they add factors like location and culture into suggestions.

8. Tech Basics

Computer vision helps find body parts and outlines using neural networks. AI learning pulls out features and finds links, and feedback helps the system keep learning and fixing itself. The system can grow, use different services, has secure access points, and has front-end platforms that are easy to use.

9. Data Ethics, Security, and Privacy

All parts of the system are designed to protect privacy. Permission is specific, all data is protected when stored and sent, and regular checks are done. Smart Size AI follows privacy rules and is open about how it makes suggestions and checks for fairness, to earn user trust.

10. Smart Size AI System Plan

The system uses a design where different parts add data, process it, use models to make guesses, and give suggestions. Security, traffic management, and backup systems make sure it's always available. Model results and retraining are part of the system.

11. Business Model and What It Offers

Smart Size AI gives value in different ways:
For shoppers: Better fit, happier customers, less wasted clothes.
For stores: Fewer returns. More sales, better image, helpful data.
For the market: Eco-friendly and social responsibility, smart stock, and standard sizing.
Data tools give clear information for stock and making products, while methods keep data private but allow general market studies.

12. How It's Done and Organized

The report has a review of what's been written, the plan for the system, tests, interviews, reviews, and pilot programs, with results in numbers and descriptions.

13. Working with Others

Working with stores, sellers, universities, and shoppers makes sure things fit the market and keeps new ideas coming. Universities help check algorithms and test pilots, and global groups push for regional changes and good deployment.

14. Testing, Checking, and Improving

Big pilot programs, trials, and feedback loops keep Smart Size AI working and reliable. Sharing results and working with researchers builds trust.

15. What It Does, How Big, and The Plan

Smart Size AI improves user experience and store profits, cuts waste, and makes a system that can grow for digital business. The plan includes AR/VR fitting, shared learning, adaptable clothes, and market-based ideas.

16. In short

Smart Size AI is a big step for ethical and scalable apparel sizing, linking AI, inclusion, business needs, and social impact. The document is a guide for users, researchers, and planners.

17. Global and Local

Smart Size AI knows that culture, language, and sizing standards affect how well sizing works. The platform adds language, measurement guides, and support. Translation makes sure things are correct, and local experts check terms. Going local means changing algorithms for body types and styles. Local stores and groups give info on what users want. Smart Size AI has sizing profiles and interfaces for different regions. Tech makes sure things run smoothly worldwide.

18. Tech Partners and System

Smart Size AI works with tech partners like cloud platforms and hardware companies. Cloud systems help with hosting, and integrations speed onboarding. Hardware helps power AR mirrors and in-store retail. Working together improves cameras and sizing. Open-source work boosts data and helps the industry.

19. Research and New Ideas

Smart Size AI spends time on research, both in-house and with universities. They look at learning, data, and privacy tech. Interests include using AI for clothes simulation, learning for advice, and shared learning for models. Studies on people drive onboarding design. The plan focuses on user privacy and comfort.

20. Summary

The clothing sizing problem is ready to change through tech. Smart Size AI is at the point of AI, design, partnerships, and sustainability, giving a plan for fashion's future. Future sections will have diagrams and case studies. Smart Size AI wants to be a solution and example for ethical tech.

LITERATURE REVIEW

1. Apparel Sizing Systems: A Look Back

Early sizing systems during the industrial age tried to categorize bodies into standard sizes for mass production. This method, focused on manufacturing, didn't take into account differences in body shapes, regional variations, or how bodies change through life. As clothing sales became global, problems appeared, such as inconsistent sizing between brands, variations in different regions, and vanity sizing, which changed what sizes like S, M, and L actually meant. As a result, people started to lose trust in size labels as a reliable way to find clothes that fit.

SmartSize AI considers this history a lesson. Instead of relying on a fixed sizing system, the platform learns from data and real-world results. Visual analysis of measurements like age, height, weight, chest, waist, and hip showed a wide, varied distribution and overlapping groups. Charts and correlation maps made it clear that a single sizing system cannot work for everyone. The answer must adapt to data, consider the brand, and personalize the experience.

2. Body Measurement in the Digital Age

Using tape measures is precise when done by experts, but it's not dependable online. SmartSize AI aims for flexibility by accepting different types of input, like direct measurements, optional fit preferences, and the possibility of adding imaging technology in the future. Industrial 3D scanners are still costly, but smartphones now allow for large-scale photogrammetry. The platform considers practical factors like lighting, pose, and agreement, which can affect the estimation quality. It is designed to combine inputs, check for errors, and rate confidence.

The project analysis found that height, weight, and torso/hip measurements were the most important factors, with demographic data giving context for preferences. The first versions concentrate on reliable and fast structured inputs, but the design allows for camera-based measurements in the future, once privacy and user experience concerns are addressed.

3. AI and Machine Learning for Size Recommendations

Traditional rule-based systems struggle with unusual cases and variations between brands. SmartSize AI tests different machine learning models: Logistic Regression, KNN, Decision Tree, Random Forest, AdaBoost, Naive Bayes, and SVMusing standard tests for accuracy, precision, recall, F1 score, confusion matrices, and breakdowns by class. SVM performed well for predicting sizes across multiple classes, while Decision Trees were useful for interpretation and quick demos. Ensemble methods provide stable baselines.

The system works by using structured inputs, preprocessing the data, using a trained model to make a prediction, and giving a recommendation with a confidence level. It also includes ways to learn from post-purchase feedback.

4. Privacy and Ethics in Sizing Tech

Sizing requires personal data. SmartSize AI focuses on privacy by collecting as little data as possible, encrypting data during transfer and storage, getting specific and revocable consent, providing clear data-retention policies, and using access controls and security plans. The system also plans to continuously check for demographic biases. The model pipelines focus on showing what factors influenced the recommendation and giving people the option to opt out. Future photographic inputs will be transparent and, when possible, preprocessed on the device.

5. Commercial Platforms and Current Industry Practices

Retailers have moved from basic charts to more advanced tools. The difficulty is ensuring consistency across different brands, materials, cuts, and inventories. SmartSize AI's modular system supports brand-specific changes, fit preferences, garment-specific guidelines (like jeans versus tops), and feedback loops that feed new data back into the training set. A Streamlit interface shows how this can be used in e-commerce, with quick input, instant results, confidence levels, and practical advice.

6. Combining Data Types for Better Analysis

The analysis phase used structured numerical and categorical data. The code normalizes numerical data and encodes categories like gender and fit preference. The design can handle different data types, including structured data now, images later, and eventually shopping behavior. Techniques for combining data include using stacked and ensemble models, checking correlations to prevent data leakage, and scoring confidence levels to handle unusual data. After deployment, the analytics layer will offer dashboards to track conversion rates, drop-off rates, and return-rate changes by size and brand.

7. Demographics, Social Factors, and Inclusivity

Preferences change based on region, culture, age, and ability. SmartSize AI aims for inclusivity by offering multiple languages, clear measurement instructions, unit conversions, and a fit preference control that has a real impact on recommendations. Future versions will add voice prompts, modes for people with low vision, and options for low-bandwidth connections. The dataset is monitored to ensure that minority body types are represented and not treated as outliers.

8. Sustainability and Environmental Concerns

Returns due to poor fit waste resources. This project links accuracy to environmental impact, such as fewer shipments, less packaging, and less excess inventory. Error heatmaps can help retailers improve their designs. Linking return rates to the recommendation system connects AI to sustainability efforts.

9. User Experience and Interface Design

Difficult interfaces reduce user completion. The Streamlit app offers a simple design with a single screen, unit options, input validation, and instant, clear recommendations. Tooltips provide guidance, preferences are easily controlled, and results show what changed when the user adjusts the fit style. This design can be added to product pages or used as a check before checkout. A/B tests are used to refine the text, order, and defaults.

10. Explaining the Algorithm and Ensuring Transparency

Users deserve to know why a size is recommended. SmartSize AI explains the main factors, gives a confidence level, and provides an alternative suggestion when confidence is low (like suggesting a larger size if a looser fit is desired). Operational dashboards track precision, recall, confusion drift by brand and garment type, and fairness. Model versions, training dates, and schema changes are documented, and human support is available for unusual cases.

11. Open-Source Platforms and Collaborative Development

This work uses and contributes to open-source methods, including scikit-learn pipelines, interpretable baselines, and research-backed model choices. The research is based on a deep learning architecture that combines user and item data, aligning with modern recommendation systems. Sharing sanitized evaluation code and synthetic datasets encourages scrutiny and speeds up improvement.

12. Regulatory and Legal Considerations

Global privacy laws require consent, purpose limits, and data-portability rights. SmartSize AI complies through configurable retention periods, access audit logs, and a consent ledger. Image-based modules use device-side preprocessing or temporary uploads to avoid storing raw images unnecessarily. A Data Protection Impact Assessment template and changelog are maintained as features are added.

13. Augmented Reality and Virtual Try-On

AR try-ons can increase engagement and reduce uncertainty, but must accurately reflect fabric and body pose. SmartSize AI uses AR to enhance, not replace, numeric fit predictions. The predictor provides size and confidence, while AR shows style and fit. Future versions will integrate cloth simulation and camera pose normalization for more realistic visuals.

14. Data Science Practices

Federated learning, secure aggregation, and automated hyperparameter tuning provide safer and better models. Model comparisons already create a culture of evidence-based decisions.

Automated retraining is triggered by drift detection, and fairness checks are included as standard steps.

15. Future Trends and Research Directions

Remaining questions include global anthropometry gaps, the need for standardization, size changes during life events, and continuous-learning policies. On-device estimation, generative fit simulation, and active learning are promising developments.

16. Feedback and Learning in Sizing Systems

The product improvement cycle involves recommendations, purchases, user feedback on fit, and retraining. Post-purchase feedback is used to adjust the model and reduce returns. Seasonal and supplier changes are tracked, and targeted data collection prompts are used for low-confidence or high-loss segments.

17. Cross-Industry Technological Impacts

Measurement tech can also improve medical garments, PPE, athletic wear, and avatar creation. Ergonomics and safety gear can use the profiling process, and consumer electronics benefit from accurate measurements.

18. Retail Adoption and Change Management

Adoption requires managing people and processes. Phased rollouts work best, starting with a few categories to prove reduced returns, then scaling up. Engineers should work with merchandising and customer experience teams. There should be clear steps for handling fit issues and version control for models and size charts. Important vendor criteria include accuracy, brand alignment, privacy, and ease of integration.

19. Limitations and Challenges

Sparse data, inaccurate self-reporting, brand drift, and low-resource environments remain difficult. SmartSize AI tries to address these through error detection, fallback logic, hybrid inputs, and documentation of known issues. Standardization of measurement schemas and APIs would improve interoperability.

Appendix A: Data & Feature Pipeline

• Inputs: Age, Height (cm), Weight (kg), Chest (cm), Waist (cm), Hip (cm), Gender, Fit Preference.

• Preprocessing: Scaling of numerical data; encoding of categorical data; trained column ordering maintained.

• Models: Logistic Regression, KNN, Decision Tree, Random Forest, Naive Bayes, AdaBoost, SVM. SVM performed best in tests, Decision Tree was clear for demos.

• Evaluation: Accuracy, Precision, Recall, F1-score, confusion matrices; diagnostics reviewed before model freeze.

• Persistence & Inference: Best model serialized. The prediction function preprocesses data and outputs a size label with confidence.

• Front-End: Streamlit UI with unit options, validation, gradient result card, helpful tips, and preference toggles.

Appendix B: Ethical & Security Controls

• Consent capture and storage; data minimization; encryption.

• Role-based access; audit logs; incident response plan.

• Model cards documenting data sources, intended use, and limitations.

• Fairness assessments and remediation plan; user opt-outs.

• For camera modules: device-side inference or temporary storage; clear prompts and delete guarantees.

Appendix C: Research

• We use content-collaborative recommenders that learn population-level information and user/item data. This helps with cold starts while maintaining personalization. The deep variant performs well on benchmarks and large datasets.

Appendix D: Deployment Notes

• Streamlit for prototyping; REST endpoint for integration.

• Blue/green deployments for model updates; A/B tests; shadow mode.

• Metrics: conversion, return rate, fit satisfaction, brand confusion.

CLIENT MEETINGS AND FEEDBACK

Identifying Reliable Size Patterns: Key Problems

Retailers have trouble spotting dependable patterns in size-related product returns. Though they collect lots of customer data, return reasons, and service feedback, the information is often unclear. Most rely on various unconnected feedback sources like store notes, online forms, call center records, and emails, which rarely use the same language. This makes it hard to pinpoint if problems come from the design, grading system, size chart, or individual customer expectations. Clients say they have data, but it's not put together well and cannot guide real action.

Imprecise Customer Feedback

Consumers rarely give exact details when they are unsatisfied with sizing. Many pick options like too small or too large, but few say if the issue is with shoulder width, chest fit, hip size, sleeve length, or the overall shape. This lack of detail leads to mistakes, and retailers have to guess when making changes. Also, personal preference affects how people view fit. One customer's relaxed fit may be another's loose or poorly structured. Retailers say this makes it harder to get consistent fixes from the feedback they receive.

Changes from Category to Category, Season to Season, and Region to Region

It's hard to stabilize fit expectations across different product groups. A size pattern in winter clothes may not apply to summer clothes. Returns for tailored items differ a lot from those for athletic wear or traditional outfits. When brands grow globally or serve different groups, the same size may mean different things. Clients point out that fashion trends change, and regional tastes vary, so even old trends don't have consistent sizing. This makes it hard to keep long-term patterns reliable, and they have to keep re-analyzing instead of building on what they've learned.

Inconsistent Measurements Between Brands and Factories

Retailers see differences between their size charts, supplier systems, and real production. The same size label doesn't always mean the same size across product lines, suppliers, or production runs. When brands work with many suppliers, these differences grow. Merchandising heads worry that even when they set standard rules, factory issues, like fabric stretch or cutting, can cause problems, which result in customer complaints.

Conflicting Information and Unclear Return Data

Another problem is the conflicting data in return reports. Often, there are just as many too-tight as too-loose complaints for the same item. Without details like body type or when the item was worn, these signals are not helpful. Merchandising teams say that such unclear issues lead to small, hesitant changes instead of big, confident changes based on data. This creates a slow cycle of trying things out rather than quick improvement.

Problems with Operations and Teamwork

There are also company issues in fixing sizing problems. Different departments have different goals for fit. Design teams might care about style, while customer support wants to lower complaints, and supply chain partners focus on factory work. Without shared data or workflows, sizing feedback is read differently across the company. Retailers say this lack of coordination causes delays, partial fixes, and repeated talks without clear solutions.

In conclusion: Better Fit Information Needed

Across client meetings, the message is the same: retailers get a lot of feedback but lack a system to turn it into sizing knowledge. The lack of consistent size language, return paperwork, connected data, and knowledge of customer groups stops them from seeing important patterns early. This leads to changes that are more reactive and scattered instead of strategic and lasting. Clients know they need a way to bring fit information together, understand customer differences, and guide choices with data to make customers happier and sizing more correct.

HARDWARE AND SOFTWARE REQUIREMENTS

Hardware requirements

The necessary hardware for implementing, testing, and showing the SmartSize AI system was planned to be very accessible, lowering technical problems for contributors, collaborators, and those who use it. The entire model creation was done in Google Colab, which works in a web browser and uses Google's cloud computing. Because of this, all computing, like preprocessing, training, tuning the model, judging it, and versioning, could be done remotely without needing strong computers. The only thing needed to participate was a device that could run a modern browser, like Chrome or Edge, and a good internet connection. Normal laptops with 4GB of RAM and dual-core processors worked well for the project, making it useful for schools, organizations, and research groups where special computers are not available. By using Colab's cloud, the project didn't need physical GPU installations or other hardware, making the whole thing easy to scale and use. This also lets people focus on thinking about the algorithms, understanding the model, and trying different things, instead of worrying about system settings or performance problems. The simple hardware also helped with showing and using the system. People could use the model through a Streamlit web page on any standard computer. No extra parts, hardware licenses, or large storage were needed. This means the solution could be used and copied easily in different schools and classrooms without needing advanced computers.

Software Requirements

The software for SmartSize AI is built for user-friendly copying and changing. All main model creation and analysis was done using Google Colab, a Python development area in a browser, powered by Jupyter notebooks. Colab works with Python libraries like NumPy for math, pandas for working with data, scikit-learn for machine learning, and Matplotlib and Seaborn for showing data. Because Colab is kept on the cloud, all important libraries were up to date and worked together, and software installation was easy. Running code in Colab let the team import needed packages using pip commands in notebook cells, so they didn't need to set things up or configure locally. For showing the model, letting people use it, and making a prototype, Streamlit was used to make interactive web apps. Streamlit allowed real-time images, user input, prediction displays, and data comments using little code. These apps were started locally and could be shared on networks without setting up remote servers or containers. Other tools used were Google Drive for saving and sharing data, and GitHub for version control and sharing the work. Using only open-source and free platforms kept the project cheap while allowing flexible scaling and new features.

Data Handling and Joining

The data set for training and judging came from Mendeley Data, making sure the data was good and right for school and work research. The data was downloaded as a CSV file and put right into the project using Google Colab. Data work was mainly done with the pandas library, allowing things like handling missing data, encoding fields, normalizing features, and joining data when needed. Because Colab works with Google Drive, team members could put shared folders into Colab sessions, helping with progress and making sure experiments could be copied. The data study included statistical analysis and images to find patterns, links, or problems that could change how the model worked. Once cleaned and structured, the data set was used for different modeling experiments, and trained model parts were later taken out for UI-level use in Streamlit. The smooth link between data storage, Colab-based work, and Streamlit app joining made a simple and clear development structure that lowered problems and allowed for trying and improving.

Ease of Use and Access

The system was made to be easy to get to for students, researchers, and groups. Because Google Colab works fully in a browser, users don't need to install Python, machine learning, or dependency managers on their computers. This lowered problems and confusion for new users. Streamlit also helped by giving a clear and interactive front-end where predictions could be tried and understood in real-time without writing code. This made it easier to show machine learning ideas, making complex modeling easy to understand and interactive for stakeholders. Also, because all the needed software was on the cloud or free to download, there were no license limits or money problems. These things made SmartSize AI good for use in schools, workshops, and research, while making sure users could be involved, no matter their technical skills or access to hardware.

Data Prep and Work Structure

The work followed a clear and repeatable structure to make the model clear and consistent. The data set was first put into Colab, then preprocessed by cleaning, normalizing, handling missing values, and encoding features. Exploratory data analysis was then done to understand distributions and links between things. After analysis, the data set was split into training and testing groups, making sure model tests showed real generalization instead of overfitting. Some machine learning algorithms were trained, like Decision Trees, K-Nearest Neighbors, Logistic Regression, and Support Vector Machines. Their work was compared using accuracy, precision, recall, F1 score, and confusion matrices. The best model was picked and taken out for use in Streamlit. The interface took user input, preprocessed it to match the trained format, and showed predictions right away. This structured way of making things made sure results were clear, correct, and easy to check, from getting data to using the interface.

Package Handling and Versioning

To keep the model able to be remade and stop problems, the project used pip-controlled library versioning in Colab notebooks. Each notebook cell for library installation had version numbers when needed. This lets the team keep the same execution on different sessions and devices. GitHub was used for code version control, allowing improvements, safe testing, and teamwork. By keeping model scripts, preprocessing, and Streamlit code in a shared place, the project ensured development was clear and made it easy to go back or add things. The notebook versioning and keeping in a repository allowed updates without hurting the system.

Security and Access Control

Security was maintained by using Google account authentication for Colab and Drive.

IMPLEMENTATION

Data Input and Setup

The initial phase of the SmartSize AI process revolved around meticulous data input and setup, which were crucial in determining the system's overall effectiveness. The project utilized a comprehensive dataset of clothing and body measurements sourced from the AWS Data Repository. This particular repository was selected for its structured organization and reliability in research, ensuring the accuracy of the information. The dataset included vital parameters such as height, weight, chest circumference, hip measurements, shoulder width, sleeve length, and waist size, along with demographic details about individuals and geographical regions relevant to sizing standards.

To facilitate seamless collaboration among the team members, the data was integrated into Google Colab through Google Drive, eliminating the need for cumbersome file transfers and ensuring that all participants were using the same data version, thus minimizing the risk of discrepancies. The dataset was subsequently transformed into pandas Data Frames for effective analysis, undergoing a thorough cleaning process to rectify any inconsistencies. Missing values were systematically addressed by filling them in with averages; however, in instances where outliers were present, median values were preferred to maintain robustness against skewed data.

Categorical variables, including gender and clothing types, were encoded using one-hot encoding, allowing for optimal compatibility with machine learning algorithms. Feature scaling was implemented using scikit-learn's StandardScaler to ensure that the dataset maintained consistency across various mathematical methods, which is essential for achieving accurate predictions. Outliers were identified and analyzed through visual tools such as boxplots and scatter plots. This process aimed to distinguish genuine variations from potential measurement errors. The thoughtful arrangement of the data was critical in priming it for subsequent analysis and mathematical modelling.

Analysing Data and Feature Engineering

Following the data cleaning phase, the next step involved delving into the dataset to uncover underlying relationships and patterns through Exploratory Data Analysis (EDA). Conducted within Google Colab, EDA employed pandas for preliminary statistics, coupled with visualization libraries like Matplotlib and Seaborn to create insightful charts. The primary objective of EDA was to comprehend the variations in measurements among individuals and identify opportunities to enhance prediction accuracy.

The analysis revealed significant relationships between specific measurements, such as the correlation between the chest-to-waist ratio and upper garment sizing, as well as the hip-to-torso ratio's influence on lower garment fit. Consequently, new derived features were engineered, and categorical options were recoded for improved machine learning performance. Additionally, a more focused EDA was undertaken to scrutinize data surrounding sizing returns. This analysis involved stratifying the data by individual profiles, clothing items, and measurement categories to uncover unusual patterns or prevalent fit issues. Seasonal trends were also analyzed, along with cultural influences affecting clothing preferences in various regions. The insights garnered from this exploratory phase significantly informed the model's architecture, feature selection, and user settings incorporated into the Streamlit interface.

Model Development and Training

The core of the machine learning endeavor was executed within Google Colab, where the team explored various methodologies to develop the most effective size prediction model. Initially, simpler models such as linear regression and decision trees were employed for baseline comparisons regarding their predictive capabilities. Subsequently, more complex ensemble methods such as Random Forests and Gradient Boosting Machines were leveraged, as these models excelled at capturing intricate relationships within the data.

Hyperparameter tuning played a pivotal role in refining model performance by adjusting critical settings, including tree depth and learning rates. Moreover, Support Vector Machines (SVM) and neural networks were introduced through TensorFlow and Keras for advanced pattern recognition tasks. Robust evaluation metrics were crucial in determining the model's efficacy; accuracy scores, confusion matrices, precision-recall balances, F1 scores, and ROC curve analyses were instrumental in identifying the best-performing models while minimizing errors. The team meticulously documented numerous model runs within the Colab notebooks, preserving detailed notes on performance outcomes. Ultimately, successful models were serialized using joblib, ensuring they could be effectively transferred to deployment environments.

Deployment with Streamlit

With a high-performing model identified, the next focus turned to developing a user-friendly interface using Streamlit. This framework facilitated the transformation of the predictive model into an accessible web application, designed for users with minimal coding proficiency. The trained models were integrated into the application through Python scripts, which not only set up the data but also transformed user inputs into formats compatible with the model.

Users could enter their measurements into the application, which would then generate size predictions along with associated confidence levels and clothing recommendations based on the input parameters. Additionally, the app incorporated visualizations from Matplotlib to elucidate the implications of the predictions. Streamlit's functionality enabled live testing of the model, fostering real-time feedback and interface improvements. The application could be accessed locally or hosted on any computer via a web browser, eliminating the need for specialized server infrastructure and thus enhancing accessibility.

Collaborative Efforts, Documentation, and Clarity

Throughout the project, effective collaboration was maintained via shared Google Drive folders, Colab notebooks, and GitHub for version control of code. Comprehensive documentation within the Colab notebooks ensured that every development stage was carefully tracked and recorded. Utilizing GitHub for version control helped keep all team members synchronized and aware of ongoing changes. The emphasis on detailed record-keeping not only assisted the team throughout the project lifecycle but also served as a valuable resource for future users. The modular architecture of the SmartSize AI system permitted the seamless addition of new data, models, or enhancements to the interface, ensuring its adaptability and growth over time.

EXPERIMENTATION AND CODE

Overview of the Implementation Process

The implementation phase of the SmartSize AI project was characterized by detailed planning, coordinated workflows, and targeted data processing. This stage translated conceptual sizing logic into a functional recommendation system. The primary objective was to develop a clothing sizing solution that would address the requirements of both shoppers seeking accurate size recommendations and retailers aiming to enhance customer satisfaction and reduce return rates. Achieving this goal required establishing a robust foundation for data handling, analysis, feature enhancement, model training, and system validation. The process ensured that measurement data were systematically organized, enabling the model to learn effectively from diverse body types and fit preferences. Emphasis was placed not only on technical precision but also on usability, maintainability, and scalability, allowing the system to adapt to evolving customer needs and industry trends.

Data Input, Integration, and Preparation

The effectiveness of the SmartSize AI model relied heavily on the quality of its data. A dataset from the AWS Data Repository was selected due to its organization and relevance to real-world clothing measurements. This dataset included essential body measurements such as height, weight, chest, waist, hips, arm length, shoulder width, and torso length. Additional attributes, age, gender, region, and fit preferences, were incorporated to help the model recognize variations across different user groups.

Data acquisition was facilitated through Google Colab, integrated with Google Drive, which enabled seamless team collaboration and eliminated file management issues. This setup allowed team members to share, modify, and review data in real time without the need for repeated uploads or synchronization. By converting the dataset into pandas DataFrames, the team could efficiently manipulate and index the data for subsequent processing.

Data cleaning was conducted meticulously to address missing values and inconsistencies. Missing numerical entries were imputed using either the mean or median, depending on the distribution, to prevent bias in the model. Categorical variables such as gender and clothing style were encoded using one-hot encoding to ensure compatibility with scikit-learn. Feature scaling was performed with StandardScaler to standardize numerical ranges, mitigating the influence of larger features on distance-based algorithms like SVM and KNN. Outliers were identified using boxplots, histograms, and scatter matrices to detect patterns that could adversely affect algorithm performance. These preparatory steps established a reliable basis for model training, evaluation, and deployment.

Exploratory Data Analysis and Insight Extraction

Exploratory Data Analysis (EDA) was undertaken to gain a comprehensive understanding of the dataset's structure, variability, and interrelationships among measurements. This analysis was instrumental in guiding feature selection and refinement prior to model development. Utilizing tools such as pandas, seaborn, and matplotlib, the team generated descriptive statistics and visualizations to examine data distributions, trends, and correlations.

Histograms revealed prevalent combinations of body measurements, offering insights into customer sizing patterns. Heatmaps were used to identify correlated features, such as the relationship between chest and shoulder width or the waist-to-hip ratio. Pair plots facilitated the exploration of multi-dimensional data patterns and the examination of measurement clusters across gender, region, or age groups.

The EDA process also highlighted the need for enhanced feature engineering to better represent body proportions. As a result, new features were created, including proportional scaling measures, ratio-based profiles, and dimension indices. These additions enriched the dataset and improved the model's ability to distinguish between sizing categories. Overall, EDA validated the data preparation approach and informed subsequent modeling decisions.

Model Training, Evaluation, and Selection

Model development was conducted within Google Colab to ensure reproducibility and facilitate collaboration. Multiple machine learning models were trained and evaluated to determine the most effective approach for predicting clothing sizes. Initial models, such as Logistic Regression and Decision Trees, established baseline performance and provided insights into feature contributions through linear relationships and hierarchical splits.

Subsequently, more advanced algorithms, including Random Forest Classifiers and Gradient Boosting, were employed to capture non-linear interactions and enhance predictive strength. Support Vector Machines (SVM) were also tested, demonstrating reliable classification performance in high-dimensional spaces. For deeper pattern recognition, neural networks were constructed using TensorFlow and Keras, enabling the model to learn complex relationships within the data.

Evaluation metrics such as accuracy, precision, recall, and F1 score were used to assess model reliability. Confusion matrices were generated to identify areas of misclassification, providing further guidance for model improvement. The final model was selected based on its performance, consistency, interpretability, and suitability for real-time application. Upon completion, the model was serialized with joblib for straightforward integration into the deployment interface.

Deployment via Streamlit User Interface

Following model development, attention shifted to accessibility and user experience. Streamlit was chosen for deployment due to its straightforward structure, intuitive widget creation, and browser-based operation without the need for a dedicated server. The user interface was designed to allow individuals to input body measurements or upload data files directly.

All input values were automatically processed through the same scaling and encoding procedures applied during training, ensuring consistency. The system then provided real-time clothing size recommendations, accompanied by explanations, confidence scores, and fit suggestions. Streamlit's visualization capabilities enabled the display of measurement distributions, prediction outcomes, and interpretative results. This interactive approach enhanced transparency and facilitated stakeholder feedback, supporting ongoing improvements to both the interface and the clarity of predictions.

Collaboration Workflow, Version Management, and Security Protocols

To maintain project continuity and reproducibility, GitHub was utilized for version control and collaborative development. Each component, from data processing scripts to Streamlit interface code, was organized within dedicated project directories, enabling modular development and efficient troubleshooting. Comprehensive documentation was maintained using markdown files, ensuring transparency throughout the project lifecycle.

Security considerations were addressed by restricting access to Google Colab and Drive through authenticated Google accounts, thereby safeguarding sensitive data and model files. This structured workflow not only supports long-term maintenance but also positions SmartSize AI for future expansion, integration with retail systems, or scaling into a larger application.

RESULTS

The development, evaluation, and deployment of the SmartSize AI model led to notable advancements in predictive size recommendation accuracy, user experience, and system interpretability when compared to traditional manual sizing methods. Early benchmark models, which utilized linear regression and basic decision tree classifiers, achieved predictive accuracies between 65% and 75%. These initial results provided a solid foundation for identifying essential data patterns and understanding feature sensitivities.

Analysis of these foundational models revealed several areas for improvement. Specifically, enhancements were needed in feature scaling, the management of regional and garment-based size differences, and the encoding of categorical variables such as gender, garment class, and fabric type. To address these issues, iterative feature engineering was introduced, incorporating ratio-based proportional metrics and demographic segmentation. These refinements directly improved model interpretability and enabled more accurate mapping of measurement variations to clothing size recommendations.

In subsequent training cycles, ensemble machine learning techniques, including Random Forest and Gradient Boosting algorithms, were adopted. These methods significantly improved the models' generalization capabilities. The ensemble models achieved accuracies above 80% for primary apparel categories, demonstrating both stability and robustness across a wide range of body measurements. They were particularly effective in modeling non-linear relationships between measurement parameters and outperformed baseline models in precision and recall for mid-range size classifications. For standard adult apparel categories such as shirts, trousers, and jackets, predictive accuracy consistently reached or exceeded 85% in cross-validation tests. However, predictions for children's apparel and specialized garments were somewhat less accurate, primarily due to greater variability in growth patterns and limited data in certain subcategories. Ongoing dataset expansion and targeted modeling strategies are expected to improve these results in future iterations.

The final model was deployed using Streamlit, resulting in an interactive, user-focused application capable of real-time size predictions. The interface enabled users to input measurement data directly, which was automatically preprocessed to align with the model's feature encoding requirements. The system provided not only recommended sizes but also confidence probabilities and explanations highlighting the influence of specific measurements on the recommendation. This transparency was essential for building user trust and enhancing perceived reliability. The application's graphical interpretation features allowed users to visually compare their body measurements with expected size distributions, further supporting interpretability and user confidence. Feedback from trial users indicated a clear preference for the SmartSize AI system over static sizing charts, particularly among those who previously faced challenges selecting appropriate sizes in traditional e-commerce settings.

Testing environments included return simulation scenarios and misfit detection protocols. Compared to manual selection or default retailer sizing tables, SmartSize AI demonstrated a measurable reduction in predicted misfit classifications. The system improved first-attempt size accuracy for most test users, reducing uncertainty-driven trial-and-error and lowering the likelihood of product returns. These outcomes highlight the practical value and commercial potential of the platform, especially for fashion retailers aiming to reduce customer dissatisfaction and minimize losses associated with size-based returns.

CONCLUSION

The SmartSize AI project has shown that accessible, open-source, and cloud-supported machine learning workflows can be effectively used to develop practical, scalable, and user-friendly sizing recommendation systems for the apparel industry. By leveraging collaborative notebook environments such as Google Colab, the development process remained efficient, reproducible, and inclusive, allowing multiple contributors to work simultaneously without the need for specialized hardware. This data-driven approach reduced dependence on intuition-based or visually estimated sizing, which has historically resulted in inconsistent outcomes and high return rates in retail.

The iterative process of feature engineering and model refinement enhanced the predictive reliability of the system, particularly when combined with advanced ensemble learning and neural network techniques. The integration of anthropometric ratios, demographic segmentation, and garment classification logic significantly improved both model accuracy and generalization. Deploying the model through Streamlit ensured that the solution was accessible in real time and understandable to non-technical users, thereby supporting strong user engagement and confidence. The interface's dynamic data visualization and explanation features were vital in making the results interpretable and preventing the model from operating as a "black box."

Despite these achievements, there remain several opportunities for further development. Future improvements may include expanding the dataset to cover a broader range of age groups, cultural fit preferences, fabric elasticity variations, and specialized apparel types such as sportswear or formalwear. Incorporating 3D body scan data or smartphone-based photogrammetric measurement could also enhance predictive accuracy. Additionally, integrating real-time feedback mechanisms within the Streamlit interface would allow users to report fit satisfaction, enabling the system to update its models continuously through adaptive learning.