Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2178~2189
e-ISSN: 2548-6861

Forecasting the Demand for Freshmen Alma Mater Jackets and Sports T-Shirts by Size Using a Hybrid Bayesian–Machine Learning Approach

Miranti Verdiana [1]*, Eko Dwi Nugroho [2]*, Leslie Anggraini [3]*, Radhinka Bagaskara [4]*
* *Teknik Informatika, Institut Teknologi Sumatera*
miranti.verdiana@if.itera.ac.id [1], eko.nugroho@if.itera.ac.id [2], leslie.anggraini@if.itera.ac.id [3], radhinka.bagaskara@if.itera.ac.id [4]

Article Info

Article history:

Received 2026-03-15
Revised 2026-04-23
Accepted 2026-05-05

Keyword:

Admissions Operations,
Bayesian Dirichlet–Multinomial,
Demand Forecasting,
Multiclass Classification,
Uniform Size Prediction.

ABSTRACT

This study addresses an operational procurement problem in university admissions, where alma mater jackets and sports T-shirts for incoming students must be ordered several months before complete size information becomes available. In the case of ITERA admissions, procurement decisions are typically made in March or April, whereas actual student size data are only gradually collected during the re-registration period from April to July. To support earlier and more reliable procurement planning, this study formulates the problem as a size-demand forecasting task covering six categories: S, M, L, XL, XXL, and XXXL. Historical data from 2015 to 2025 were analyzed, with reliable size records concentrated in the 2019–2025 period. The main novelty of this study lies in formulating freshman uniform procurement as a staged forecasting problem that follows the actual admissions workflow. Specifically, the study proposes a hybrid framework that combines: (i) a time-weighted Bayesian Dirichlet–Multinomial model for early-stage aggregate forecasting when current-year size data are not yet available, and (ii) a CatBoostClassifier-based multiclass machine learning model for prediction updates when student attributes become available. Model performance was evaluated using an expanding-window rolling/forward chaining scheme with a one-year forecasting horizon. In addition to conventional historical baselines, the study also included Simple Exponential Smoothing (SES) as a time-series benchmark. Performance was assessed using cross-entropy for size-distribution accuracy, MAE/size and WAPE for quantity prediction, and stockout/overstock for operational impact. The results show that the previous-year proportion remains a strong baseline, while the Bayesian model provides competitive performance and yields posterior uncertainty estimates that are useful for determining safety-oriented order quantities. The statistical analysis further confirms that gender is the most influential predictor of size, while study program, admission track, and province provide complementary but weaker signals. The findings indicate that the proposed framework can support more adaptive and evidence-based procurement planning, reduce the risk of size shortages and excess inventory, and provide a transferable forecasting workflow that may be adapted to other institutions after local recalibration.

I. INTRODUCTION

The procurement of alma mater jackets and sports T-shirts for freshmen is a time-critical operational process because both items must be ready before the official orientation and distribution schedule. In practice, however, complete size information is not yet available when procurement decisions must be made. In the case of ITERA admissions, uniform orders are generally placed in March or April, whereas actual student size data are collected gradually during the re-registration period from April to July. This time lag between data availability and procurement planning creates substantial uncertainty in determining how many units should be ordered for each size category. From an operations and supply chain perspective, inaccurate demand estimation may lead to stockout for specific sizes or overstock, both of which negatively affect service quality, inventory efficiency, and procurement cost [1]–[3].

The problem addressed in this study is therefore not limited to forecasting the total number of freshmen. More specifically, it concerns forecasting the demand for each size category S, M, L, XL, XXL, and XXXL because procurement decisions are made on a size-by-size basis rather than in aggregate. Moreover, the size distribution of freshmen is not necessarily stable over time. Variations in the composition of incoming students, such as gender, study program, admission track, and province of origin, may influence the proportion of each size category. As a result, a reliable forecasting framework must be able to learn from historical size patterns while also incorporating available student attributes when such information becomes available. In modern predictive modeling practice, robust forecasting performance depends on appropriate preprocessing, categorical-feature handling, and evaluation procedures that preserve the original structure of the prediction problem [4].

From the state-of-the-art perspective, demand forecasting in the fashion and apparel sector has increasingly adopted artificial intelligence and machine learning approaches to improve performance under uncertain and dynamic conditions [5]. Research in retail demand forecasting likewise shows that supervised learning methods can support operational decisions when demand is affected by changing contextual patterns [6]. However, most of these studies focus on forecasting sales volume, whereas the present study deals with a different problem: forecasting the distribution of size categories for institutional uniforms. This makes the problem closer to compositional forecasting, where the objective is to estimate the proportion of several mutually exclusive categories. In this context, Dirichlet-based models have been used to forecast compositional data over time [7], while Dirichlet regression has been applied to proportional outcomes in forecasting settings [8]. At the count level, Dirichlet–Multinomial formulations are also suitable for modeling categorical count data while accounting for uncertainty [9]. When student-level attributes are available, the problem can be extended to a multiclass classification task using categorical predictors. CatBoost is particularly relevant because it is designed to handle categorical variables effectively and to reduce bias caused by target leakage through ordered boosting [10], [11]. In addition, studies in the textile and fashion domain show that machine learning can support decisions related to body measurement and size prediction problems [12]. To ensure that model assessment remains valid under temporal conditions, this study applies rolling/forward chaining evaluation, which is consistent with recommended practice for time-dependent prediction tasks [13].

Based on this context, the objective of this study is to develop a forecasting framework for estimating the demand for freshmen alma mater jackets and sports T-shirts by size for a target admission year. Operationally, the target output is defined as in Equation 1.

$$\widehat{D_{t,k}} = \widehat{N_t} \times \widehat{p_{t,k}}\ , \qquad \text{Equation 1}$$

Where $\widehat{N_t}$ is the predicted number of incoming students and $\widehat{p_{t,k}}$ is the predicted proportion of size k. To achieve this objective, the study proposes a hybrid approach consisting of a time-weighted Bayesian Dirichlet–Multinomial model for early-stage aggregate forecasting and a CatBoost Classifier-based multiclass machine learning model for prediction updates when student attributes become available during re-registration.

The novelty of this study is threefold. First, unlike prior work that mainly focuses on retail sales forecasting or general apparel demand, this study formulates uniform procurement in university admissions as a multiclass size-demand forecasting problem with direct operational outputs in the form of order quantities by size. Second, the proposed hybrid framework is designed as a staged decision-support mechanism that mirrors the actual procurement workflow: a Bayesian aggregate forecast is used when current-year size data are not yet available, and a machine-learning-based update is applied when partial current-year attributes become available. Third, the study explicitly links predictive modeling with operational procurement decisions by quantifying uncertainty, stockout, overstock, and safety-oriented order plans, thereby extending the contribution from pure prediction to practical supply planning.

This study is guided by four research questions. First, how can the demand for freshmen alma mater jackets and sports T-shirts be forecasted by size category before complete current-year size data is available? Second, to what extent do student attributes such as gender, study program, admission track, and province improve forecasting performance compared with historical-proportion and time-series baselines? Third, can a hybrid approach combining Bayesian compositional forecasting and machine-learning-based prediction updates provide better support for procurement planning than conventional baseline methods? Fourth, how should the resulting forecasts be translated into practical procurement decisions in order to reduce the risk of stockout and overstock for specific size categories?

This study is limited to forecasting the size demand for freshmen alma mater jackets and sports T-shirts under the assumption that both items use the same size label for each student. The analysis uses institutional admissions data from 2015 to 2025, but size-distribution modeling is focused on the years with sufficiently reliable size records, particularly 2019–2025. The predictive variables are restricted to admissions attributes available in the dataset, namely gender, study program, admission track, and province of origin. Anthropometric variables such as height and weight are not included because they are not systematically available in the current records. The findings are therefore expected to be transferable to other institutions only after recalibration to local cohort characteristics, size distributions, and data-collection practices.

Accordingly, the contribution of this study is not limited to predictive accuracy, but extends to the design of an uncertainty-aware and operationally interpretable procurement decision framework for staged admissions data.

II. METHOD

This study uses a quantitative, data-driven forecasting approach to predict the size needs of new students' alma mater jackets and sports jerseys. The method's design aligns with the operational context of PMB, where ordering decisions are made in March/April when individual size information is incomplete, while still meeting the requirements of scientific evaluation on temporal (annual) data.

A. Research Design

This study adopts a quantitative, data-driven forecasting approach to estimate the demand for freshmen alma mater jackets and sports T-shirts by size. The method is designed to reflect the actual operational setting in which procurement decisions must be made several months before complete current-year size information becomes available. Accordingly, the forecasting framework is divided into two stages. The first stage produces an early aggregate forecast based on historical size distributions, while the second stage updates the forecast when student-level attributes become available during the re-registration period. Since the data are annual and sequential, model performance is evaluated using rolling/forward chaining, so that each test year is forecasted only from preceding years and no future information is used during training [13].

B. Dataset and Variables

The dataset consists of freshmen admissions records from 2015 to 2025. The available variables include student entry year, uniform size label, gender, study program, admission track, and province of origin. In this study, the size of the alma mater jackets and the sports T-shirt is assumed to be identical for each student, so that one size label represents the demand for both items. Student identifiers such as name and student number are excluded from the predictive model and are used only for data integration and validation, in line with standard machine learning practice [4].

A data-quality audit was conducted before model development. Table I shows that the completeness of size records is highly uneven across years. The years 2015–2017 contain no valid size labels, and 2018 contains only one valid size record. In contrast, the years 2019–2025 provide highly reliable size data, with completeness ranging from 99.72% to 100.00%. This sharp improvement justifies the use of 2019–2025 as the reliable period for modeling size distribution. The choice of this period is therefore not arbitrary, but is based on explicit data-quality evidence. Furthermore, Table II and Table III show that this reliable period also contains sufficient variation in cohort size and size composition to support temporal forecasting analysis.

For the update-stage model, the predictor variables are gender, study program, admission track, and province. All of these variables are categorical and are therefore handled using a classifier that supports categorical inputs directly. In this study, CatBoostClassifier was selected because it can process categorical features without extensive one-hot encoding and includes ordered boosting to reduce prediction shift caused by target leakage [10], [11].

TABLE I
COMPLETENESS OF SIZE RECORDS BY YEAR (2015–2025)

Years	Total Data	Filled Size	Blank Size	% Filled
2015	397	0	397	0
2016	1298	0	1298	0
2017	1598	0	1598	0
2018	2677	1	2676	0.04
2019	3664	3664	0	100
2020	4566	4565	1	99.98
2021	4698	4685	13	99.72
2022	4873	4860	13	99.73
2023	4434	4434	0	100
2024	4484	4484	0	100
2025	4935	4935	0	100

C. Data PreProcessing

Several preprocessing steps are applied before modeling. First, size labels are standardized into six categories: □, □, □, *XL*, *XXL*, and *XXXL*. Second, invalid or missing province values are mapped to an Unknown category to avoid unnecessary row deletion. Third, admission-track labels are normalized across years to reduce category drift caused by naming changes. Fourth, duplicate records are resolved so that each student contributes only one final record for a given entry year. These preprocessing steps are intended to improve consistency and reduce noise in categorical modeling, which is especially important for algorithms that rely on stable feature representations [4], [10].

D. Forecasting Target Formulation

The operational forecasting target is defined as the vector of procurement quantities for each size category in year □ which can be seen in Equation 2:

$$\widehat{D_t} = [\widehat{D_{t,S}} \times \widehat{D_{t,M}} \times \widehat{D_{t,L}} \times \widehat{D_{t,XL}} \times \widehat{D_{t,XXL}} \times \widehat{D_{t,XXXL}}\ ]. \quad \text{Equation 2}$$

This vector is estimated through the decomposition on Equation 3:

$$\widehat{D_{t,k}} = \widehat{N_t} \times \widehat{p_{t,k}}\ , \quad \text{Equation 3}$$

where $\widehat{N_t}$ is the predicted total number of freshmen in year t and $\widehat{p_{t,k}}$ ais the predicted proportion of size □ in the same year. This formulation is consistent with demand forecasting in inventory planning, where total demand and category- level demand composition jointly determine the procurement quantity [1]–[3].

E. Baseline Methods

Three baseline methods and one additional time-series benchmark are used for comparison. The first baseline is Last-Year Proportion, where the predicted size distribution for year t is assumed to be identical to the observed distribution in year

$t-1$, i.e., $\widehat{p_t} = p_{t-1}$. The second baseline is the Three-Year Moving Average, which averages the observed distributions of the three most recent years which can be seen in Equation 4.

$$\widehat{p_t} = \frac{1}{3}(p_{t-1} + p_{t-2} + p_{t-3}). \qquad \text{Equation 4}$$

The third baseline is Pooled Proportion, which estimates the size distribution from the aggregate counts across all training years which can be seen in Equation 5.

$$\widehat{P_{t,k}} = \frac{\sum_{y\,\epsilon\,\top} \quad c_{y,k}}{\sum_{y\,\epsilon\,\top} \quad N_{y,}}. \qquad \text{Equation 5}$$

In addition to these historical baselines, this study also includes Simple Exponential Smoothing (SES) as a time-series benchmark. SES is applied to the annual size proportions with a smoothing parameter α selected through temporal backtesting on the training window. The SES benchmark is included to provide a stronger comparison against a standard time-series approach that emphasizes recent observations while still smoothing short-term fluctuations. This additional benchmark is important because the forecasting problem exhibits inter-year drift, as later shown in Table III and Fig. 1. These baselines and benchmarks are simple, interpretable, and widely used as transparent reference points in forecasting studies [1], [5].

F. Proposed Hybrid Forecasting Approach

1) *Stage-1 Early Forecasting: Time-Weighted Bayesian Dirichlet–Multinomial:* In the early procurement stage, current-year student-level size information is not yet available. Therefore, the first stage focuses on forecasting the aggregate size distribution using a Bayesian Dirichlet–Multinomial model. Let c_t is [$c_{t,S}$, $c_{t,M}$, $c_{t,L}$, $c_{t,XL}$, $c_{t,XXL}$, $c_{t,XXXL}$] denote the vector of observed size counts in year t, where $c_{t,k}$ is the number of students assigned to size category k. The total cohort size in year t is defined as on Equation 6.

$$N_t = \sum_k \quad c_{t,k}. \qquad \text{Equation 6}$$

The observed count vector is assumed to follow a multinomial distribution conditional on the latent size-proportion vector $\mathbf{p}_t$ likelihood is defined as $\mathbf{c}_t \mid \mathbf{p}_t, N_t \sim$ Multinomial(N_t,$\mathbf{p}_t$), where $\mathbf{p}_t$ is [$\mathbf{p}_{t,S}$, $\mathbf{p}_{t,M}$, $\mathbf{p}_{t,L}$, $\mathbf{p}_{t,XL}$, $\mathbf{p}_{t,XXL}$, $\mathbf{p}_{t,XXXL}$] denotes the latent size-proportion vector for cohort t. A Dirichlet prior is then assigned to the proportion vector $\mathbf{p}_t \sim$ Dirichlet($\boldsymbol{\alpha}_0$), where $\boldsymbol{\alpha}_0$ is the prior concentration vector. In this study, a weakly informative symmetric prior is used so that each size category receives positive prior mass.

To emphasize more recent cohorts, the Bayesian updating process incorporates temporal weighting through a forgetting factor □ ∈ (0, 1]. For a training window □ whose most recent year is denoted by T, the posterior concentration vector is computed as which can be seen in Equation 7:

$$\alpha^* = \alpha_0 + \sum_{y \in T} \quad \lambda^{(T-y)}\, c_y. \qquad \text{Equation 7}$$

This formulation assigns larger weights to more recent cohorts and smaller weights to older cohorts. The posterior mean of the proportion for size category k is then obtained as on Equation 8.

$$\widehat{p_{t,k}} = \frac{\alpha_k^*}{\sum_j \quad \alpha_k^*}. \qquad \text{Equation 8}$$

This Bayesian formulation is suitable for the present problem for two reasons. First, it models size demand directly as a compositional quantity across mutually exclusive categories. Second, it provides stable estimates for low-frequency categories such as XXXL while also yielding posterior uncertainty that can later be translated into safety-oriented procurement quantities [7]–[9].

2) *Stage-2 Forecast Updating: Attribute-Based Multiclass Machine Learning:* When student-level attributes become available during the re-registration period, the forecasting process is refined through a multiclass classification model. For each student i, let x_i denote the attribute vector consisting of gender, study program, admission track, and province, and let y_i denote the observed size category. The update-stage model estimates the conditional probability $Pr(y_i = k \mid x_i)$, $k \in \{S,M,L,XL,XXL,XXXL\}$.

In this study, the machine-learning algorithm used in the update stage is CatBoostClassifier. This algorithm was selected because it handles categorical predictors directly, reduces the need for extensive one-hot encoding, and applies ordered boosting to reduce target leakage and prediction shift [10], [11]. After class probabilities are obtained for all students in the target cohort, the expected procurement quantity for size category k is computed by aggregating the individual probabilities on Equation 9:

$$\widehat{D_{t,k}^{ML}} = \sum_{i \in cohort(t)} \quad Pr(y_i = k \mid x_i)\,. \qquad \text{Equation 9}$$

This aggregation converts individual-level probabilistic predictions into cohort-level demand estimates, making the output directly relevant to operational procurement planning. [4].

3) *Integration of the Two Stages:* The proposed framework is hybrid because it combines the strengths of Bayesian aggregate forecasting and machine-learning-based updating in a staged manner. In the first stage, the Bayesian Dirichlet–Multinomial model is used to generate an early aggregate forecast when current-year individual size data are not yet available. In the second stage, once partial current-year records and student attributes become available, the forecast is updated by combining the observed counts and the predicted counts for the remaining students. Formally, the updated procurement estimate is defined as on Equation 10.

$$\widehat{D_t} = C_t^{obs} + \widehat{D_t^{rem}}. \quad \text{Equation 10}$$

Here, C_t^{obs} denotes the observed count vector for students whose sizes have already been recorded, while $\widehat{D_t^{rem}}$ denotes the predicted demand vector for the remaining students whose sizes are still unknown. In practical terms, the Bayesian model supports the initial procurement decision, whereas the CatBoostClassifier-based update model refines that decision as additional admissions information becomes available. This staged integration is the main operational novelty of the proposed hybrid forecasting framework.

G. Evaluation Scheme and Metrics

The models are evaluated using an expanding-window rolling/forward chaining scheme across four test years: 2022, 2023, 2024, and 2025. For each fold, all years prior to the test year are used for training, and the forecasting horizon is one year ahead. Specifically, the folds are 2019–2021 → 2022, 2019–2022 → 2023, 2019–2023 → 2024, and 2019–2024 → 2025. This configuration reflects the actual decision setting in which procurement for year t must be determined using information available before year t [13]. Performance is measured at three complementary levels. First, the quality of the predicted size distribution is evaluated using cross-entropy on Equation 11.

$$CE(t) = -\sum_k \quad p_{t.k} log(\widehat{p_{t,k}}), \quad \text{Equation 11}$$

Second, the quality of the predicted procurement quantities is evaluated using MAE per size and WAPE which can be seen in Equation 12 and Equation 13.

$$MAE(t) = \frac{1}{K}\sum_k \quad |D_{t,k} - \widehat{D_{t,k}}|, \quad \text{Equation 12}$$

$$WAPE(t) = \frac{\sum_k \quad |D_{t,k}-\widehat{D_{t,k}}|}{\sum_k \quad D_{t,k}}. \quad \text{Equation 13}$$

Third, the operational consequences of forecasting error are evaluated through stockout and overstock which can be seen in Equation 14 and Equation 15.

$$Stockout = \sum_k \quad max\left(0, D_{t,k} - \widehat{D_{t,k}}\right), \quad \text{Equation 14}$$

$$Overstock = \sum_k \quad max\left(0, \widehat{D_{t,k}} - D_{t,k}\right). \quad \text{Equation 15}$$

For the Bayesian model, uncertainty is additionally quantified using posterior credible intervals for the size proportions and predictive intervals for the implied order quantities. These uncertainty summaries are then used to derive safety-oriented order plans based on predictive quantiles, such as Q95.

H. Implementation and Reproducibility

All experiments are implemented using consistent temporal splits without random shuffling across years. Preprocessing rules, category mappings, and model settings are kept fixed across folds. Hyperparameters for the Bayesian model, especially the forgetting factor are selected through temporal backtesting within the training window rather than using the test year. Likewise, the machine learning stage is trained only on the training years of each fold. This keeps the experimental design reproducible and consistent with time-aware forecasting practice [4], [13].

III. Results and Discussion

A. Data Exploration and Data Quality

1) *Completeness of Size Records by Year:* Before modeling, the completeness of size labels was examined to determine whether the historical data were sufficiently reliable for learning stable size-distribution patterns. As shown in Table I, the completeness of size records is highly uneven across years. In 2015–2017, size information is entirely unavailable, while in 2018 only one valid size record is present. By contrast, the period 2019–2025 shows very high completeness, ranging from 99.72% to 100.00%. This has an important methodological implication: although the institutional dataset spans 2015–2025, the forecasting of size demand should be trained primarily on 2019–2025, since earlier years do not provide a sufficient basis for learning the distribution of size categories. Therefore, 2015–2018 are treated as a data limitation for size forecasting rather than as a valid training period.

2) *Size Distribution and Inter-Year Changes:* The overall size distribution for the 2019–2025 cohorts is summarized in Table II. The results show that size L is the dominant category, followed by M and XL. The year-by-year distribution is presented in Table III, while the temporal trend is illustrated in Fig. 1. As shown in Table III and Fig. 1, the distribution is not fully stationary. The proportion of L remains high in 2019–2022 but gradually declines in the later years. Meanwhile, XL rises sharply and reaches its peak in 2023 before decreasing again in 2024–2025. A different pattern is observed for M, whose proportion declines until 2022–2023 and then rebounds in 2025. Most importantly, XXXL begins to appear in a non-negligible proportion only from 2023 onward. This pattern confirms that the forecasting problem is not simply a matter of repeating a fixed historical average; rather, the model must account for temporal drift in the composition of size demand.

TABLE II
Overall Size Distribution (2019–2025)

Size	N (People)	Percentage (%)
S	1241	3.92
M	8758	27.69
L	13401	42.37
XL	6398	20.23
XXL	1585	5.01
XXXL	244	0.77
Total	31627	100

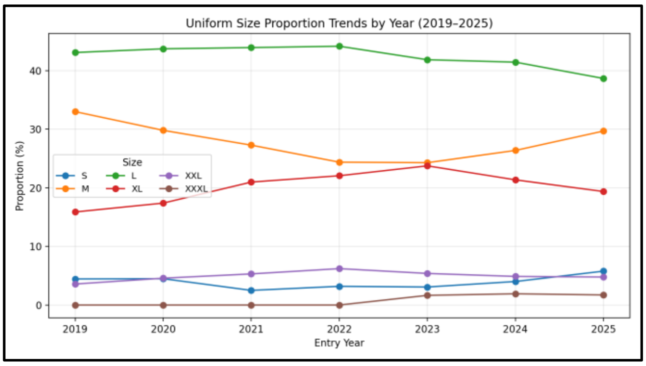

Fig. 1. Size trends 2019 - 2025

TABLE III
SIZE DISTRIBUTION BY YEAR (PERCENT, 2019–2025)

Years	N (People)	S	M	L	XL	XXL	XXXL
2019	3664	4.45	33	43.09	15.88	3.58	0
2020	4565	4.49	29.81	43.72	17.39	4.58	0
2021	4685	2.5	27.28	43.93	20.98	5.31	0
2022	4860	3.19	24.38	44.16	22.06	6.21	0
2023	4434	3.07	24.29	41.86	23.75	5.39	1.65
2024	4484	4.01	26.38	41.44	21.36	4.88	1.92
2025	4935	5.78	29.69	38.66	19.37	4.78	1.72

3) *Feature Signal, Relationship Between Student Attributes and Size:* The clearest signal is observed for gender, as shown in Table IV and Fig. 2. Among male students, the distribution is concentrated more strongly in larger sizes, especially L, XL, and XXL, whereas female students show substantially higher proportions of S and M. This suggests that gender is the strongest individual predictor of uniform size.

TABLE IV
SIZE DISTRIBUTION BY GENDER (PERCENT, 2019–2025)

Gender	N (People)	S	M	L	XL	XXL	XXXL
M	15805	0.87	18.03	47.14	26.42	6.43	1.1
F	15822	6.97	37.34	37.61	14.05	3.59	0.44

The relative strength of association between each student attribute and uniform size is summarized in Table V. Gender has the largest association value, substantially exceeding study program, province, and admission track. Thus, while the latter variables may still contribute complementary information, gender clearly provides the strongest predictive signal.

To complement the descriptive comparison, a x^2 test of independence was conducted for each categorical feature against the size category. The results indicate that all four attributes are significantly associated with size ($p < 0.001$), but with different effect magnitudes. Gender shows the strongest association ($x^2 = 2753.31$, df = 5, $p < 0.001$, Cramér's V = 0.295), while study program ($x^2 = 1013.60$, df = 205, $p < 0.001$, V = 0.080), admission track ($x^2 = 524.79$, df = 65, $p < 0.001$, V = 0.058), and province ($x^2 = 416.11$, df = 160, $p < 0.001$, V = 0.051) have weaker but still statistically significant associations. These results confirm that gender is the dominant predictor of size, whereas the other features function as secondary refinement variables rather than primary drivers.

TABLE V
ASSOCIATION STRENGTH (CRAMÉR'S V) BETWEEN FEATURES AND SIZE

Feature	Cramer's V vs Size	N	Number of Categories
Gender	0.29	31627	2
Study program	0.08	31627	42
Province	0.05	31627	33
Entry path	0.03	31627	7

Variation across study programs is presented in Table VI and Fig. 3. The share of large sizes (XL+) varies meaningfully across the Top-10 programs by record count, indicating that program composition may influence the demand for larger uniforms. The corresponding pattern by normalized admission track is shown in Table VII and Fig. 4, while the provincial comparison is reported in Table VIII and Fig. 5. Although these effects are weaker than the gender effect, they suggest that program, track, and province can still provide useful refinement signals for update-stage forecasting.

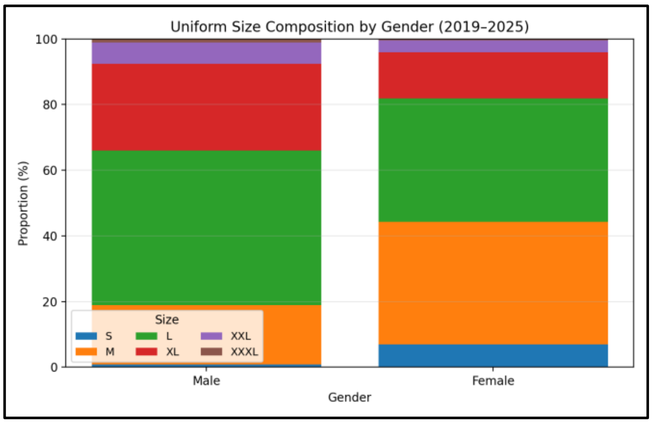

Fig. 2. Size composition by gender

TABLE VI
SIZE DISTRIBUTION BY STUDY PROGRAM (TOP-10 BY N, PERCENT)

Study Program	N (People)	S	M	L	XL	XXL	XXXL	XL+ (%)
Teknik Informatika	1603	2.62	24.2	43.17	23.58	5.36	1.06	30
Teknik Sipil	1555	2.19	25.4	45.47	20.51	5.79	0.64	26.94
Perencanaan Wilayah dan Kota	1524	5.25	32.87	39.83	16.27	5.25	0.52	22.04
Teknik Pertambangan	1339	2.09	23.75	45.78	22.4	5.38	0.6	28.38
Farmasi	1279	5.86	40.89	33.54	15.09	4.07	0.55	19.71
Teknik Geofisika	1215	3.7	27.16	43.13	19.59	5.35	1.07	26.01
Teknik Industri	1175	3.57	30.47	44.6	16.94	4.17	0.26	21.37
Teknik Elektro	1164	1.55	20.96	45.02	25.86	5.76	0.86	32.48
Teknik Lingkungan	1105	5.61	30.59	39.91	18.46	4.8	0.63	23.89
Arsitektur	1055	4.55	30.9	42.94	17.06	4.27	0.28	21.61

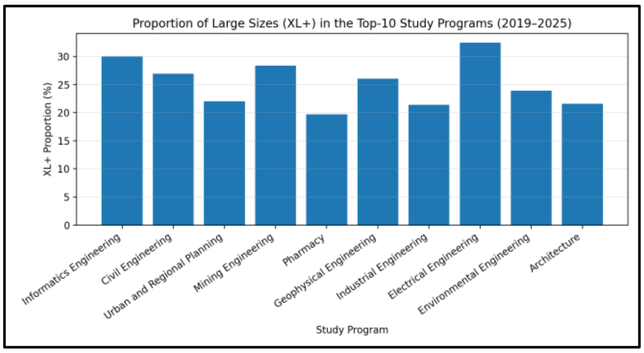

Fig. 3. XL+ by study program

TABLE VII
SIZE DISTRIBUTION BY ADMISSION TRACK (NORMALIZED, PERCENT)

Entry Path	N (People)	S	M	L	XL	XXL	XXXL	XL+ (%)
SN- Test	17031	4	27.06	41.93	20.85	5.2	0.95	27
SN- Achievement	12412	3.83	28.79	43.38	19.18	4.41	0.41	24
Independent	1357	3.39	26.09	39.79	21.96	7.37	1.4	30.73
Achievement	651	4.92	27.5	38.86	20.28	6.76	1.69	28.73
USM / USMK	128	2.34	24.22	48.44	20.31	4.69	0	25
Other	44	4.55	27.27	43.18	22.73	2.27	0	25
Golden Ticket	4	0	0	50	0	25	25	50

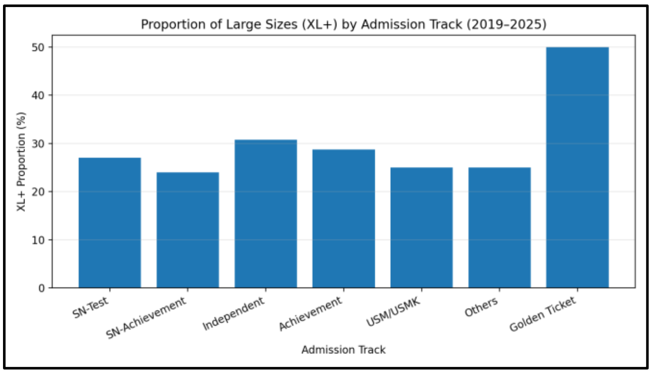

Fig. 4. XL+ by admission track

TABLE VIII
SIZE DISTRIBUTION BY PROVINCE (TOP-10 BY N, PERCENT)

Province	N (People)	S	M	L	XL	XXL	XXXL	XL+ (%)
Prov. Lampung	14893	4.23	28.71	42.6	19.22	4.63	0.6	24.45
Prov. Sumatera Utara	5086	3.38	27.59	46.01	19.31	3.22	0.49	23.02
Prov. Jawa Barat	2650	3.47	25.32	41.06	22.11	7.06	0.98	30.15
Prov. Sumatera Selatan	2312	4.11	27.98	40.83	19.98	5.88	1.21	27.07
Prov. Sumatera Barat	1808	2.38	26.11	41.48	23.78	4.98	1.27	30.03
Prov. D.K.I. Jakarta	1541	4.48	27.97	37.44	20.7	7.98	1.43	30.11
Prov. Banten	1506	3.78	25.56	40.57	22.38	6.84	0.86	30.08
Prov. Riau	661	3.18	22.54	41.91	26.78	4.08	1.51	32.37
Prov. Jambi	398	6.03	32.41	35.43	17.84	7.04	1.26	26.14
Prov. Bengkulu	225	5.33	27.56	39.11	23.56	4	0.44	28

B. *Experimental Design and Evaluation Setup*

The empirical results of the evaluation setup described in Chapter II are reported in Tables IX–XIX. The rolling/forward chaining scheme ensures that each test year is predicted only from preceding years, which allows the results to be interpreted as genuine out-of-sample forecasting performance.

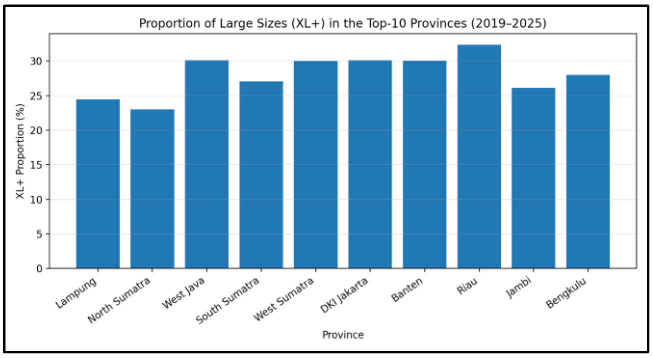

Fig. 5. XL+ by province

C. *Baseline Results*

1) *Baseline Performance on Size Distribution Forecasting:* The fold-level baseline results are reported in Table IX, while the mean performance across folds is summarized in Table X. As shown in Table X, the Last-Year Proportion baseline achieves the lowest mean WAPE and the most competitive cross-entropy. This indicates that the most recent historical distribution is generally a stronger predictor than either a three-year moving average or a pooled long-run average. In other words, the size distribution exhibits enough year-to-year drift that older years can dilute the relevance of recent patterns.

TABLE IX
BASELINE RESULTS PER FOLD (2022–2025)

Method	Test Year	N	Cross-Entropy	MAE / Size	WAPE	Stockout (pcs)	Overstock (pcs)
Baseline-LastYear	2022	4860	1.3245	47	0.058	141	141
Baseline-MA3	2022	4860	1.3343	101.67	0.1255	305	305
Baseline-Pooled	2022	4860	1.3331	96.67	0.1193	290	290
Baseline-LastYear	2023	4434	1.7543	49.33	0.0668	148	148
Baseline-MA3	2023	4434	1.7577	78	0.1055	234	234
Baseline-Pooled	2023	4434	1.7612	95.67	0.1295	287	287
Baseline-LastYear	2024	4484	1.4026	49.67	0.0665	149	149
Baseline-MA3	2024	4484	1.4122	52.67	0.0705	158	158
Baseline-Pooled	2024	4484	1.4181	49	0.0656	147	147
Baseline-LastYear	2025	4935	1.4333	83.33	0.1013	250	250
Baseline-MA3	2025	4935	1.4422	124	0.1508	372	372
Baseline-Pooled	2025	4935	1.4423	93.67	0.1139	281	281

TABLE X
MEAN BASELINE RESULTS (2022–2025)

Method	Test Year	N	Cross-Entropy	MAE / Size	WAPE	Stockout (pcs)	Overstock (pcs)
Baseline-LastYear	2023	4678.25	1.4787	57.3325	0.0731	172	172
Baseline-MA3	2023	4678.25	1.4866	89.085	0.1131	267.25	267.25
Baseline-Pooled	2023	4678.25	1.4887	83.7525	0.1071	251.25	251.25

To strengthen the baseline comparison, Simple Exponential Smoothing (SES) was also evaluated as an additional time-series benchmark. As shown in Table X-A, SES achieved a mean cross-entropy of 1.4790 and a mean WAPE of 0.0773. These results are slightly weaker than the Last-Year Proportion baseline (mean WAPE = 0.0731), but still clearly stronger than the MA3 and pooled baselines. This comparison suggests that the forecasting problem is highly sensitive to recent cohort information, and that aggressively recency-oriented benchmarks remain difficult to outperform.

TABLE X-A
SIMPLE EXPONENTIAL SMOOTHING (SES) RESULTS (2022–2025)

Method	Test Year	N	Cross-Entropy	MAE / Size	WAPE
SES	2022	4860	1.3247	51.67	0.0638
SES	2023	4434	1.7543	51.33	0.0695
SES	2024	4484	1.4029	51.33	0.0687
SES	2025	4935	1.4343	88.33	0.1074
Mean	-	-	1.4790	60.67	0.0773

2) *Baseline Performance on Procurement Quantities and Error Patterns:* The best-performing baseline for each test year is presented in Table XI. As shown in Table XI, the dominant forecasting errors are concentrated in sizes M, L, and XL, because these categories account for the largest share of the total demand. For 2022, the best baseline overestimates M and underestimates XL. In 2023, it overestimates L while underestimating XL. In 2025, the strongest baseline underestimates M and overestimates L, which is consistent with the change in the observed distribution shown earlier in Table III and Fig. 1. These findings are consistent with the aggregate performance reported in Table X. At the same time, Table XI also highlights a practical weakness for minority classes: in 2024, the best baseline underestimates XXXL by a substantial amount, which is operationally critical because shortages in very large sizes are difficult to absorb.

TABLE XI
BEST BASELINE PER TEST YEAR + LARGEST ERRORS

Test Year	Best Baseline (WAPE)	WAPE	Cross-Entropy	Biggest Mistake #1	Biggest Mistake #2
2022	Baseline-LastYear	0.058	1.3245	M (+141)	XL (-52)
2023	Baseline-LastYear	0.0668	1.7543	L (+102)	XL (-75)
2024	Baseline-Pooled	0.0656	1.4181	L (+87)	XXXL (-71)
2025	Baseline-LastYear	0.1013	1.4333	M (-163)	L (+137)

D. Results of the Bayesian Model

1) *Performance Compared with Baselines:* The fold-level performance of the Bayesian model is reported in Table XII, and its mean performance is summarized in Table XIII. Overall, the Bayesian Dirichlet–Multinomial model provides competitive results, although it does not outperform the strongest baseline in every evaluation fold. The best performance is observed in 2024, where the Bayesian model achieves WAPE = 0.0540 and MAE/size = 40.33, both of which are lower than the best baseline for that year. This indicates that the time-weighted posterior can balance recency and stability effectively when the target-year size composition remains reasonably aligned with the recent historical pattern.

However, in 2022, 2023, and 2025, the Bayesian model remains weaker than the strongest recency-based baseline in procurement-oriented metrics. This pattern suggests that although the Bayesian model is able to smooth fluctuations and produce coherent probabilistic estimates, it may lag when the target year exhibits a sharper distribution shift than expected from the weighted historical data. In other words, the model is strongest when recent historical structure remains informative, but less effective when the cohort-level size composition changes more abruptly from one year to the next.

TABLE XII
BAYESIAN RESULTS PER FOLD (2022–2025)

Method	Test Year	N	λ*	Cross-Entropy	MAE / Size	WAPE	Stockout (pcs)	Overstock (pcs)
Bayesian-DM (Weighted)	2022	4860	0.6	1.329	77.33	0.0955	232	232
Bayesian-DM (Weighted)	2023	4434	0.6	1.4536	72	0.0974	216	216
Bayesian-DM (Weighted)	2024	4484	0.6	1.4075	40.33	0.054	121	121
Bayesian-DM (Weighted)	2025	4935	0.6	1.4384	106	0.1289	318	318

TABLE XIII
MEAN BAYESIAN RESULTS (2022–2025)

Method	Cross-Entropy	MAE / Size	WAPE	Stockout (pcs)	Overstock (pcs)
Bayesian-DM (Weighted)	1.4071	73.915	0.094	221.75	221.75

2) *Uncertainty Quantification and Stability on Minority Sizes:* A key advantage of the Bayesian model is that it provides full posterior uncertainty rather than only point estimates. As reported in Table XVIII, the predicted 2026 size proportions are accompanied by 95% credible intervals. For example, the predicted proportion for size L is 40.76% with a 95% credible interval of 39.86%–41.67%, while the predicted proportion for size XXXL is 1.44% with a 95% credible interval of 1.23%–1.66%. These interval estimates make the forecasting output more informative for procurement planning because they explicitly quantify the uncertainty around each size category.

The uncertainty perspective is particularly important for minority sizes such as XXXL. As shown earlier in Table III and Fig. 1, XXXL only begins to appear as a meaningful category in the later cohorts. Accordingly, the Bayesian model responds by gradually assigning nonzero posterior mass to that category once it becomes observable in the historical record. This makes the model more stable than purely static historical-proportion approaches when a new minority size begins to emerge. Nevertheless, the model remains conservative for newly emerging large-size categories. In the early years where XXXL is absent or nearly absent from the training window, the posterior mean for XXXL remains small, which leads to underestimation. Therefore, the Bayesian model improves stability, but it cannot completely eliminate the structural difficulty of forecasting a low-frequency category with limited historical support.

3) *Sensitivity of the Forgetting Factor and Safety-Stock Implications*: The tuned forgetting factor selected for each evaluation fold is reported in Table XIV. As shown in Table XIV, the same value of \(\lambda = 0.6\) is repeatedly selected across all folds. This result indicates that recent years consistently carry greater predictive relevance than older observations, which is consistent with the temporal drift observed in Fig. 1. The repeated selection of the same \(\lambda\) value also supports the internal consistency of the time-weighted Bayesian specification. At the same time, the fold-level results in Table XII suggest that even an appropriately tuned forgetting factor cannot fully eliminate lag when the target-year shift is abrupt. Therefore, the Bayesian model should be interpreted as an early-stage forecasting tool that balances recency with probabilistic stability rather than as a universally dominant predictor.

From an operational perspective, the most important contribution of the Bayesian model lies in its direct support for safety-oriented procurement planning. As shown in Table XIX, the comparison between posterior mean and Q95 order quantities provides an explicit basis for safety-stock determination. The additional buffer can be defined as which can be seen in Equation 16.

$$SS_k = Q95_k - \mu_k, \quad \text{Equation 16}$$

where SS_k denotes the safety stock for size category k, $Q95_k$ denotes the 95th percentile predicted quantity, and μ_k denotes the mean predicted quantity. Under the example scenario with N = 5000, the resulting buffers are +30 for S, +64 for M, +68 for L, +56 for XL, +31 for XXL, and +17 for XXXL. In supply-chain terms, this buffer serves as a service-oriented safeguard against forecast uncertainty. The policy is particularly relevant for XL, XXL, and XXXL because under-ordering these categories is operationally more costly than holding a limited amount of additional stock. Hence, the Bayesian model contributes not only by forecasting expected demand, but also by translating predictive uncertainty into concrete safety-stock recommendations for procurement planning.

TABLE XIV
TUNED LAMBDA PER FOLD (INTERNAL BACKTESTING)

Test Year	λ*
2022	0.6
2023	0.6
2024	0.6
2025	0.6

E. Results of the Update-Stage Model (CatBoost)

1) *Fold-Level Performance Compared with the Best Baseline:* The fold-level performance of the CatBoost update-stage model is presented in Table XV, and the average results are summarized in Table XVI. Compared with the strongest baseline results in Table XI, the CatBoost model improves performance only in 2024. In 2022, 2023, and 2025, the model produces larger procurement-oriented errors than the best baseline. This means that attribute-based forecasting in its initial configuration is not yet strong enough to consistently outperform a powerful recency-based benchmark.

TABLE XV
CATBOOST (FAST) RESULTS PER FOLD (2022–2025)

Method	Test Year	N	Cross-Entropy	MAE / Size	WAPE	Stock out (pcs)	Over stock (pcs)	Macro - F1	Unique Classes in Train
CatBoost - Multi class (Fast)	2022	4860	1.33	114.6	0.14	344	344	0.18	5
CatBoost - Multi class (Fast)	2023	4434	1.76	104	0.14	312	312	0.14	5
CatBoost - Multi class (Fast)	2024	4484	1.41	43.67	0.05	131	131	0.14	6
CatBoost - Multi class (Fast)	2025	4935	1.44	95.67	0.11	287	287	0.14	6

TABLE XVI
MEAN CATBOOST (FAST) RESULTS (2022–2025)

Method	Cross-Entropy	MAE / Size	WAPE	Stock out (pcs)	Over stock (pcs)	Macro-F1
CatBoost - Multi class (Fast)	1.4895	89.50	0.1142	268.5	268.5	0.1544

2) *Why the Update-Stage Model Does Not Yet Beat the Baseline:* This result should be interpreted together with the feature analysis in Tables IV–VIII. Although Table V shows that gender has the strongest association with size, the overall forecasting results in Tables XV and XVI indicate that attribute-based prediction alone is not yet sufficient to consistently outperform the strongest temporal baseline. One reason is that the strongest baseline implicitly captures temporal drift, whereas the CatBoost model is trained as a cross-sectional *Pr(y|x)* classifier without an explicit time component. Another reason is that the available attributes do not fully explain the year-to-year variation in size

composition. Finally, minority classes such as XXXL remain difficult to learn when historical support is limited.

3) *What Should Be Improved to Support the Claimed Contribution:* A more realistic assessment of the update-stage contribution should be based on partial- availability evaluation, as summarized in Table XIX. The current direct-cohort results in Tables XV and XVI therefore represent only an initial benchmark rather than the final intended use case of the update-stage model. In the actual admissions workflow, the CatBoostClassifier model is not designed to replace the early-stage Bayesian forecast entirely, but to improve the forecast once partial current-year information has been collected. Under this staged setting, the role of machine learning is to reduce residual uncertainty among students whose sizes have not yet been recorded.

4) *To strengthen the update-stage contribution, future work should include three refinements:* First, a formal ablation study should be conducted to quantify the incremental contribution of gender, study program, admission track, and province. Second, deeper hyperparameter tuning and probability calibration should be applied to improve multiclass demand aggregation. Third, the update model should be evaluated under realistic partial-observation scenarios that mimic the actual sequence of admissions data collection from April to July. These refinements would provide a fairer and more operationally meaningful assessment of the hybrid framework.

TABLE XVII
PARTIAL AVAILABILITY EFFECT (BAYESIAN) - MEAN WAPE VS OBSERVED COVERAGE

Test Year	0% (Models Only)	25% Observed	50% Observed	75% Observed
2022	0.0955	0.0728	0.0498	0.0259
2023	0.0974	0.0741	0.0496	0.0261
2024	0.054	0.0417	0.0275	0.0163
2025	0.1289	0.0964	0.0649	0.0324

F. Operational Procurement Analysis

1) *Stockout and overstock risk by size:* The procurement implications of the three model groups can be compared using Tables X, XIII, and XVI. As shown in Tables X, XIII, and XVI, the strongest baseline still yields the lowest average stockout and overstock. However, aggregate stockout/overstock does not fully capture the operational risk. Shortages in XXXL or XXL are often more problematic than shortages in dominant sizes because redistribution or substitution is much less feasible. Therefore, procurement policy should not rely solely on average WAPE; it should also consider the asymmetric operational cost of underestimating large-size categories.

2) *Order Plan for 2026 and Procurement Recommendation:* The predicted 2026 size proportions are reported in Table XVII and visualized in Fig. 6. The corresponding procurement recommendations under multiple cohort-size scenarios are presented in Table XVIII. For practical procurement, the most defensible policy is to use the posterior mean as the base order and the Q95 quantity as a safety-oriented upper reference. This leads to a "mean + targeted buffer" strategy, in which the additional quantity is defined as which can be seen in Equation 17.

$$Safety\ Stock_k = Q95_k - Mean_k \qquad \text{Equation 17}$$

Under the N = 5000 scenario, the recommended buffers are +30 for S, +64 for M, +68 for L, +56 for XL, +31 for XXL, and +17 for XXXL. In supply-chain terms, this buffer can be interpreted as a service-oriented safeguard against forecast uncertainty. The policy is particularly relevant for sizes XL, XXL, and XXXL because underordering these categories is operationally more costly than holding a small additional inventory. Therefore, the practical implication of the Bayesian uncertainty model is not merely statistical; it directly supports safety-stock determination and more transparent procurement planning.

TABLE XVIII
PREDICTED 2026 SIZE PROPORTIONS (BAYESIAN DM, □ = 0.6)

Size	Posterior Mean (%)	CI95 (%)
S	4.51	4.14-4.90
M	27.58	26.76-28.43
L	40.76	39.86-41.67
XL	20.67	19.93-21.43
XXL	5.03	4.65-5.45
XXXL	1.44	1.23-1.66

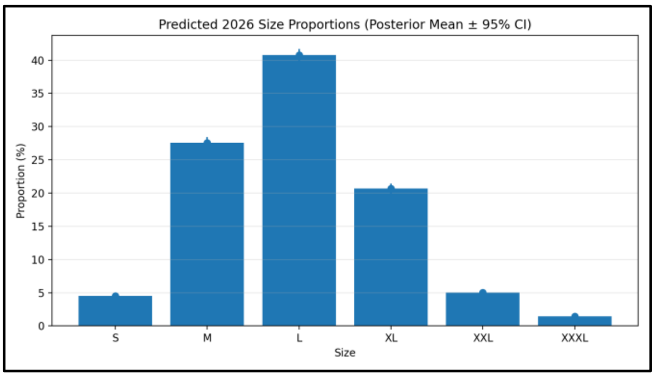

Fig. 6. Predicted 2026 size proportions with credible intervals

TABLE XIX
ORDER PLAN 2026 (MEAN VS Q95) FOR MULTIPLE N SCENARIOS

Scenario N	Size	Mean (pcs)	Q95 (pcs)	Buffer (Q95-Mean)
N=5000 (example)	S	225	255	30
N=5000 (example)	M	1380	1444	64
N=5000 (example)	L	2037	2105	68
N=5000 (example)	XL	1034	1090	56
N=5000 (example)	XXL	252	283	31
N=5000 (example)	XXXL	72	89	17
N=5000 (example)	Total	5000	5000	0

Scenario N	Size	Mean (pcs)	Q95 (pcs)	Buffer (Q95-Mean)
N=4935 (naive)	S	222	251	29
N=4935 (naive)	M	1362	1425	63
N=4935 (naive)	L	2011	2078	67
N=4935 (naive)	XL	1021	1079	58
N=4935 (naive)	XXL	248	279	31
N=4935 (naive)	XXXL	71	87	16
N=4935 (naive)	Total	4935	4935	0
N=5004 (trend)	S	226	255	29
N=5004 (trend)	M	1380	1444	64
N=5004 (trend)	L	2040	2108	68
N=5004 (trend)	XL	1035	1090	55
N=5004 (trend)	XXL	252	283	31
N=5004 (trend)	XXXL	72	89	17
N=5004 (trend)	Total	5004	5004	0

G. Discussion

1) *Interpretation of the main findings:* The main empirical pattern emerging from Tables X, XIII, and XVI is that temporal baselines remain very strong, while the Bayesian model adds value through uncertainty-aware forecasting. The update-stage model confirms that student attributes contain useful predictive signals, but that signal alone is not yet strong enough to consistently surpass a recency-based baseline under direct cohort forecasting.

2) *This study has several limitations:* First, the effective training window for size forecasting is restricted to 2019–2025 because earlier years do not contain sufficiently reliable size labels. Second, the forecasting framework assumes that the alma mater jacket and sports T-shirt use the same size label for each student; this assumption simplifies modeling, but it may not hold in other institutional settings. Third, the available predictor set is limited to admissions attributes and does not include anthropometric variables such as height or weight, which could plausibly improve individual-level size prediction. Fourth, the machine learning update model is evaluated in an initial configuration and has not yet undergone full hyperparameter optimization or calibration. Finally, the transferability of the model to other universities remains conditional on local recalibration, because cohort composition, regional body-size patterns, admissions practices, and size-reporting behavior may differ substantially across institutions.

3) *Implications for future work:* The partial-availability results in Table XIX suggest that staged updating is a promising direction for future implementation. Further work should therefore focus on formal staged evaluation, deeper tuning and calibration of the update-stage model, and the integration of richer predictors if they become available.

Although the proposed workflow is designed using ITERA admissions data, its structure is potentially transferable to other institutions that face similar procurement timing constraints. However, direct reuse of the estimated parameters is not recommended because size distributions may vary across universities, regions, and student populations. Therefore, the transferable component of this study is the staged forecasting workflow itself namely, data-quality screening, early Bayesian aggregate forecasting, and attribute-based forecast updating whereas the estimated distributions and safety stock levels should be recalibrated using local institutional data.

Accordingly, the contribution of this study is not limited to predictive accuracy, but extends to the design of an uncertainty-aware and operationally interpretable procurement decision framework for staged admissions data.

H. Suggestions

For practical implementation, it is recommended that procurement planning follow a two-stage strategy. In the first stage, an early order should be prepared using the Bayesian forecast together with a targeted safety stock based on Q95 predictive quantities, especially for XL, XXL, and XXXL. In the second stage, the order plan should be updated progressively during the re-registration period by combining observed size entries with predicted demand for students whose size is still missing. This strategy is expected to reduce both stockout risk and excess inventory while remaining consistent with the actual timing of admissions operations.

For future research, three directions are recommended. First, the update-stage model should be evaluated under a realistic partial-availability scenario that reflects the actual arrival pattern of current-year admissions data. Second, a formal ablation study and more extensive hyperparameter tuning should be conducted to better quantify the contribution of each predictor. Third, if additional variables such as height, weight, or other anthropometric information become available, they should be incorporated to improve minority-size prediction and strengthen overall forecasting accuracy.

IV. Conclusion

This study developed a forecasting framework for estimating the demand for freshmen alma mater jackets and sports T-shirts by size in order to support procurement planning before complete current-year size data become available. The data audit showed that the years 2019–2025 provide the only reliable basis for modeling the size distribution, while earlier years are unsuitable because of severe missingness. The exploratory analysis further showed that size L remains the dominant category, but the proportions of M, L, and XL change across cohorts, while XXXL emerged as a meaningful minority category in recent years. The empirical results show that recency-based forecasting remains highly competitive. The Last-Year Proportion baseline is the strongest overall benchmark, and SES also performs strongly as an additional time-series comparator. The time-weighted Bayesian Dirichlet Multinomial model does not dominate all baselines, but it contributes an important practical advantage by producing posterior uncertainty, credible intervals, and safety-oriented order quantities. The CatBoostClassifier-based update-stage model, in its initial form, does not yet consistently outperform the strongest recency-based baseline under direct cohort forecasting; however, its intended value lies in staged updating when partial current-year information becomes available. Overall, the main contribution of this study lies in proposing a staged hybrid forecasting framework that matches the admissions-procurement workflow and links predictive modeling to operational decisions. Beyond point forecasting, the framework provides a quantitative basis for uncertainty-aware procurement, safety-stock determination, and progressive forecast updating during the re-registration period.

REFERENCES

[1] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021.

[2] S. Chopra, Supply Chain Management: Strategy, Planning, and Operation, 7th ed., Global ed. Pearson, 2019.

[3] J. Heizer, B. Render, and C. Munson, Operations Management: Sustainability and Supply Chain Management, 13th ed. Pearson, 2020.

[4] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[5] K. Swaminathan and R. Venkitasubramony, "Demand forecasting for fashion products: A systematic review," Int. J. Forecast., vol. 40, no. 1, pp. 247–267, 2024, doi: 10.1016/j.ijforecast.2023.02.005.

[6] J. Huber and H. Stuckenschmidt, "Daily retail demand forecasting using machine learning with emphasis on calendric special days," Int. J. Forecast., vol. 36, no. 4, pp. 1420–1438, 2020, doi: 10.1016/j.ijforecast.2020.02.005.

[7] H. Katz, K. T. Brusch, and R. E. Weiss, "A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times," Int. J. Forecast., vol. 40, no. 4, pp. 1556–1567, 2024, doi: 10.1016/j.ijforecast.2024.01.004.

[8] C. Hanretty, "Forecasting multiparty by-elections using Dirichlet regression," Int. J. Forecast., vol. 37, no. 4, pp. 1666–1676, 2021, doi: 10.1016/j.ijforecast.2021.03.007.

[9] W. D. Wadsworth, R. Argiento, M. Guindani, J. Galloway-Peña, S. A. Shelburne, and M. Vannucci, "An integrative Bayesian Dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data," BMC Bioinformatics, vol. 18, Art. no. 94, 2017, doi: 10.1186/s12859-017-1516-0.

[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in Advances in Neural Information Processing Systems 31 (NeurIPS 2018), 2018, pp. 6639–6649.

[11] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," arXiv preprint arXiv:1810.11363, 2018, doi: 10.48550/arXiv.1810.11363.

[12] P. Meyer, B. Birregah, P. Beauseroy, E. Grall, and A. Lauxerrois, "Missing body measurements prediction in fashion industry: a comparative approach," Fash. Text., vol. 10, Art. no. 37, 2023, doi: 10.1186/s40691-023-00357-5.

[13] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," Comput. Stat. Data Anal., vol. 120, pp. 70–83, 2018, doi: 10.1016/j.csda.2017.11.003.