Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Using Autoencoders for Anomaly and Drift Detection in Linguistic Segmentation on Product Review Platforms and Recommendation Systems

Raghu K Para

Independent Researcher, Windsor, Ontario, Canada

Product review platforms are vital for the consumer technology and digital commerce ecosystem, offering insights into buying preferences, satisfaction levels, and trends. They support personalization, product improvements, and inventory management. However, their effectiveness can be undermined by irregularities in data, such as fraudulent reviews and shifts in consumer language. This paper explores the use of autoencoders—an unsupervised learning architecture—for detecting anomalies and concept drift in customer feedback. Building on research in anomaly detection, concept drift adaptation, and autoencoder architecture, we propose a robust framework for accurately identifying anomalies and monitoring drift. Using the Amazon Product Reviews Dataset, we validate our approach, achieving high precision in anomaly detection and reliable drift monitoring over time.
We provide visualizations, pseudo-code for reproducibility, and practical deployment suggestions. Our findings demonstrate that combining linguistic segmentation with unsupervised modeling enhances system robustness, ensuring recommendation engines remain trustworthy and relevant amidst evolving language and malicious manipulation.

1. Introduction

In the contemporary digital marketplace, consumer technology and digital commerce platforms are integral to the users' decision-making processes. They not only host a broad range of products or services but also surface consumer feedback through reviews, ratings, and textual comments. These user-generated content streams offer a wealth of linguistic data that, when analyzed and segmented systematically, can reveal nuanced consumer sentiments, evolving preferences, and brand perception shifts over time. Automated recommendation systems leverage these insights to guide consumers toward products of interest, improve search relevance or accuracy, and inform sellers about market gaps and opportunities.

Despite their utility and value, the integrity and reliability of these recommendation engines are often threatened by several adversarial and environmental factors. For instance, fraudulent or unwarranted reviews—generated by bots or unscrupulous sellers aiming to manipulate product perceptions—pose a significant challenge for businesses. Such anomalies can harshly influence and degrade the quality of recommendations, hamper consumer trust, and ultimately hurt the platform's reputation (Xu et al., 2020). Additionally, concept drift, defined as changes in the underlying distribution of data over time, can largely impact the accuracy of models that rely on static and siloed assumptions. As consumer language and interests evolve—introduction of new jargon, internet slang, seasonal buying trends, novel product or service categories—the models trained on historical data may fail to dynamically adapt, resulting in obsolete or irrelevant recommendations (Gama et al., 2014; Zhang et al., 2020).

This paper focuses on addressing these twin challenges—anomaly detection and drift adaptation—through the lens of autoencoders, an unsupervised neural network architecture that learns efficient data representations in latent space. Autoencoders are adept at capturing intricate structures in complex, high-dimensional spaces, such as those found in text-based datasets. Their ability to reconstruct normal data patterns, while failing to reconstruct anomalous or previously unseen patterns with the same precision and fidelity, makes them ideal for outlier detection tasks (Vincent et al., 2010). Moreover, by continuously monitoring reconstruction errors over time, these models can also identify when shifts in data distributions occur naturally, thereby enabling effective drift detection and dynamic adaptation strategies.

Within this context, linguistic segmentation—dividing text into coherent, meaningful units such as topics, sentiments, or key phrases—further enhances the interpretability and exactness of our approach. Segmenting reviews into thematic units allows the autoencoder to focus on more contextually homogeneous patterns, improving both anomaly detection and drift monitoring. The Amazon Product Reviews Dataset serves as our case study, offering a substantially rich and diverse corpus for model training and evaluation. We present metrics such as precision, recall, and F1-score for anomaly detection and rely on measures like KL divergence for drift detection to quantify performance improvements.

The remainder of this paper is organized as follows: Section 2 deeply surveys the relevant literature on anomaly detection, concept drift, and the role of autoencoders in handling unstructured textual data. Section 3 details the methodology, including data preprocessing steps, the autoencoder architecture, and the linguistic segmentation approach. Pseudo-code that entails a sequence of steps to execute the test is provided to guide practitioners. Section 4 presents the experimental setup followed by the results, highlighting improvements over baseline approaches. In Section 5, we discuss the practical implications of our findings, address key challenges, and suggest avenues for future work. Finally, Section 6 concludes by emphasizing the significance of robust anomaly and drift detection frameworks in maintaining reliable and adaptive recommendation systems.

2. BACKGROUND AND RELATED WORK

A. Anomaly Detection Using Autoencoders

Anomalies, also referred to as outliers, represent data points that deviate significantly from the expected patterns within a dataset (Hawkins, 1980). In the context of product reviews, anomalies may present as spam content, abnormally short or long reviews, syntactically awkward or ambiguous text, or suspiciously repetitive language aimed at tampering the platform's ranking system. Traditional anomaly detection methods often rely on statistical thresholds or supervised models trained on labeled anomalies. However, labeling anomalies is inherently challenging and computationally expensive, given their rarity and the complexity of defining "normal" behavior in a dynamic ecosystem.

Autoencoders, which offer a compelling unsupervised alternative, are designed to minimize reconstruction error for "normal" instances, and learn compressed representations of input data (Vincent et al., 2010). When presented with anomalous input, the model's reconstruction error tends to increase due to a poor fit to learned or trained representations. Studies have demonstrated this approach's efficacy in text domains (Xu et al., 2020), cybersecurity (Aggarwal, 2017), and spam detection (Sakurada & Yairi, 2014). Other complementary methods, such as Isolation Forests (Liu et al., 2008), can be integrated as well to enhance robustness, particularly in high-dimensional spaces. Yet, autoencoders remain popular due to their neural network-based flexibility, scalability, and ability to capture nonlinear structures.

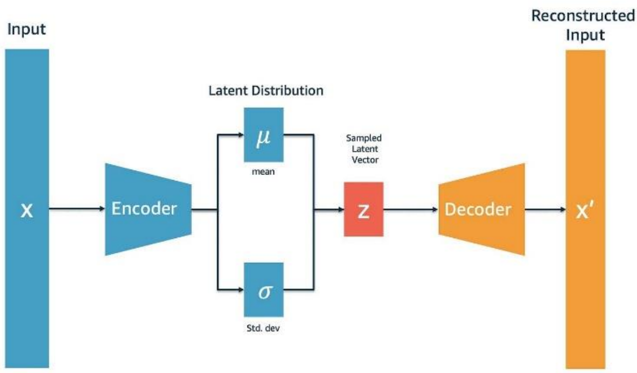

Fig 1. Anomaly Detection Using Autoencoders

B. Drift Detection and Concept Adaptation

Concept drift is a critical phenomenon that occurs when the statistical properties of a data stream change over time, causing predictive models that depend on historical data to become less accurate (Gama et al., 2014). In the domain of product reviews, drift can stem from various sources: new product or service categories or brands entering the market, shifts in consumer demographics, evolving consumer language usage (e.g., trending slang terms, new descriptors for quality or aesthetics), and seasonal trends (Zhang et al., 2020).

Early methods of drift detection focused on statistical tests, window-based monitoring of performance metrics, or adaptive retraining schedules. However, these approaches may not scale well when dealing with large and unstructured textual data. Autoencoders present an opportunity to detect drift by monitoring changes in reconstruction error distributions over a span of time. As new linguistic patterns emerge or existing ones dwindle, the autoencoder experiences difficulty in accurately reconstructing inputs that deviate from its trained representation. By continuously evaluating reconstruction errors across time windows, drift can be identified to trigger follow-up adaptation measures. This framework aligns closely with established drift detection methodologies in streaming data analysis (Snyder et al., 2019).

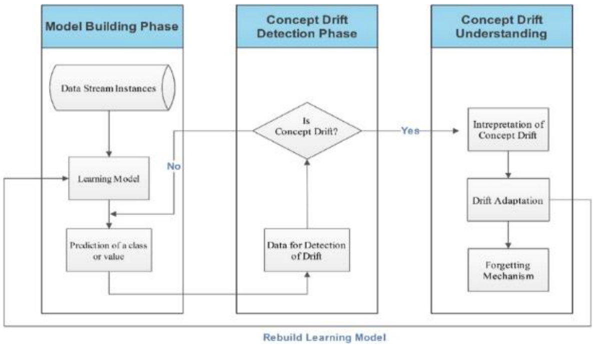

Fig 2. Main framework for concept drift detection and adaptation.

C. Linguistic Segmentation for Enhanced Interpretability

The field of natural language processing (NLP) offers numerous techniques for transforming raw text into more structured, meaningful representations. Linguistic segmentation encompasses a variety of tasks, including tokenization, stemming, lemmatization, named entity recognition (NER), and topic modeling. For this research, we focus particularly on topic modeling via Latent Dirichlet Allocation (LDA) and related semantic clustering methods (Bishop, 2006). This segmentation process is not merely a preprocessing step but it also adds a layer of interpretability and finesse to anomaly and drift detection.

When topics are well-defined, anomalies can be more easily attributed to specific contexts—such as a sudden surge of unwarranted, unnatural text within a certain product or service category. Similarly, drift detection becomes more meaningful when it can specify changes within the thematic spectrum of the review corpus. Segmenting text into coherent topics ensures that the autoencoder's latent space matches well with meaningful linguistic structures. Prior research has shown that combining segmentation with advanced machine learning architectures enhances both accuracy and precision in interpretive tasks (Goodfellow et al., 2016).

D. Relevance of Unsupervised Methods in Dynamic Environments

The digital commerce sector thrives on dynamism, with product or service offerings, consumer trends, and competitive landscapes evolving rapidly. Unsupervised learning methods like autoencoders are well-suited to this environment, as they do not rely on large amounts of labeled data. Instead, they leverage the inherent structure of the data to discover the usual patterns. In constantly shifting linguistic environments, this flexibility is invaluable. As new products or services emerge and language evolves, an unsupervised model can adapt more readily than a supervised equivalent that would require continuous re-labeling and computational retraining.

In summary, the literature suggests that anomaly detection and drift adaptation are crucial components of maintaining robust recommendation or personalization systems. Autoencoders, supported by linguistic segmentation, present a promising approach to tackling these challenges simultaneously. This synergy not only preserves systemic performance but also contributes to higher-quality consumer experiences and journeys and platform integrity.

3. METHODOLOGY

A. Dataset and Data Sources

Our experiments utilize the Amazon Product Reviews Dataset, which encompasses roughly 1 million reviews spread across a diverse array of product categories, including electronics, home appliances, apparel, and personal care items are sold. Each review includes textual data, timestamps, product identifiers, and rating data. While ratings are not directly used in the anomaly detection process, they can provide contextual understanding during analysis.

The primary preprocessing pipeline involves the following steps (Bishop, 2006):

- Tokenization: Splitting each review into individual tokens (words or subwords).
- Stopword Removal: Excluding common words that do not contribute meaningfully to semantic understanding (e.g., “the,” “and,” “or”).
- Stemming and Lemmatization: Converting words to their root forms to control sparsity and ensure consistent use of vocabulary.
- TF-IDF Vectorization: Transforming processed text into numerical feature vectors that capture the importance of each term relative to the corpus in the latent space. TF-IDF stands for Term Frequency–Inverse Document Frequency and helps signify the keywords that carry more weight within the corpus.

B. Linguistic Segmentation with Topic Modeling

After obtaining TF-IDF vectors, we apply Latent Dirichlet Allocation (LDA) to discover latent topics. LDA clusters words into coherent themes and associates each review with a probability distribution over these topics. The number of topics is chosen empirically, balancing interpretability with specificity. For example, a choice of 25 topics may provide enough granularity to distinguish fine-grained themes such as “sustainable products,” or “premium electronics”).

Segmenting this corpus into topics allows us to run the autoencoder on these thematically consistent segments. By training separate autoencoders for each topically defined smaller set or by incorporating topic distributions as additional input features, we achieve granular control over anomaly detection and drift observation. This approach limits the heterogeneity of input distributions, making the autoencoder’s reconstruction error a more sensitive indicator of deviations.

C. Autoencoder Model Architecture

We design a stacked autoencoder to handle high-dimensional TF-IDF input vectors. A typical input vector might have tens of thousands of dimensions, depending on the size of the vocabulary. The autoencoder architecture is composed of:

- Input Layer: Matches the dimensionality of the TF-IDF feature space.
- Encoder Layers: A series of fully connected layers with decreasing dimensionality, compressing input vectors into a latent representation. Non-linear activations (e.g., ReLU) are employed to capture complex relationships.
- Latent Layer: The bottleneck of the network, representing the compressed knowledge of the input distribution.
- Decoder Layers: A symmetric mirror of the encoder, aiming to reconstruct the original input from the latent representation.

We utilize the Adam optimizer for parameter updates and a mean squared error (MSE) loss function to measure reconstruction constancy. Training is conducted over 50 epochs with a batch size of 64, as inspired by Vincent et al. (2010). Hyperparameter tuning, such as choosing the learning rate or the number of hidden units, is empirical and performed via grid search or Bayesian optimization on a validation set.

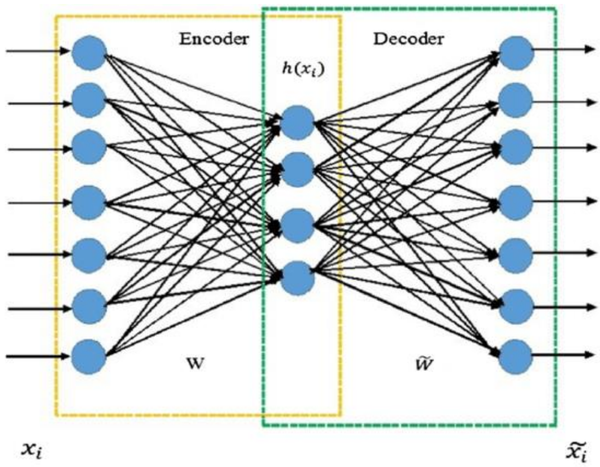

Fig 3. Autoencoder Architecture

D. Pseudo-code for Anomaly and Drift Detection

Below is a pseudo-code outline integrating autoencoder-based anomaly and drift detection, inspired by Sakurada & Yairi (2014):

1) Preprocess the dataset:

  a) Tokenize and remove stopwords from each review.

  b) Apply stemming/lemmatization.

  c) Compute TF-IDF vectors for all reviews.

2) Apply topic modeling (LDA):

a) Extract let’s say K topics.

b) Assign each review a topic distribution vector.

3) Select a modeling strategy:

Option A: Train one autoencoder per topic cluster.

Option B: Train a single autoencoder on the entire corpus, with topic distributions as features.

4) Define Autoencoder Architecture:

input_dim = dimension_of_TFIDF

encoder_layers = [Dense(...), Activation(...), ...]

latent_layer = Dense(latent_dim, activation='relu')

decoder_layers = [Dense(...), Activation(...), ...]

model = combine(encoder_layers, latent_layer, decoder_layers)

model.compile(optimizer='adam', loss='mse')

5) Train the autoencoder on norma lbaseline data:

model.fit(train_data, train_data, epochs=50, batch_size=64, validation_split=0.1)

6) Determine an anomaly threshold:

reconstruction_errors = model.predict(train_data) - train_data

threshold = mean(reconstruction_errors) + c * std(reconstruction_errors)

[the choice of c is a scalar chosen by means of validation or domain knowledge]

7) Anomaly Detection:

For each new review vector r:

error = |model.predict(r) - r|

If error > threshold:

Mark r as anomalous.

8) Drift Detection:

Segment data by time windows (e.g., monthly batches).

For each time window:

Compute average reconstruction error E_t.

Track E_t over time:

If E_t deviates significantly from historical baseline (e.g., via KL divergence or statistical test):

Indicate concept drift and adapt model or threshold.

E. Evaluation Metrics

1) Anomaly Detection Metrics: Precision, recall, and F1-score measure the model's ability to correctly identify anomalies without labeling too many normal instances as anomalies.

2) Drift Detection Metrics: We utilize Kullback-Leibler (KL) divergence (Kingma & Welling, 2013; Zhang et al., 2020) to quantify differences in the distribution of reconstruction errors over time. A significant shift indicates that the model's previously learned distribution is no longer matched with current data.

Table 1. Evaluation Metrics for Anomaly and Drift

Metric	Description	Purpose
Precision	The proportion of true positive results among all positive predictions.	Measures the accuracy of anomaly detection.
Recall	The proportion of true positives among all actual positives.	Assesses the model's ability to detect all anomalies.
F1-Score	The harmonic mean of precision and recall.	Balances precision and recall in performance evaluation.
KL Divergence	A measure of how one probability distribution diverges from a second expected distribution.	Quantifies changes in reconstruction error distributions over time for drift detection.
Mean Squared Error (MSE)	The average of the squares of the errors between predicted and actual values.	Used to evaluate reconstruction quality in autoencoders.
AUC-ROC	Area Under the Receiver Operating Characteristic Curve.	Evaluates the trade-off between true positive rate and false positive rate.

4. EXPERIMENTS AND RESULTS

A. Experimental Setup

We split the Amazon reviews dataset into an 80% training set and a 20% testing set. The training set covers a baseline period representing historical user behavior, while the testing set spans a more recent period, allowing for the assessment of the drift. Within the testing period, we synthetically inject anomalies by introducing reviews with random word insertions or repetitive spam phrases, ensuring a controlled environment for measuring anomaly detection performance. Additionally, we select a subset of products known to have undergone linguistic shifts—such as a rising popularity in sustainable products—to evaluate drift detection.

Hyperparameters, such as the number of topics ($K = 50$), latent dimensionality (e.g., 100 units), and the threshold multiplier c for anomaly detection, are chosen based on validation experiments. We also test variations in model complexity (e.g., adding more encoder-decoder layers) to confirm that our chosen architecture is neither overfitting nor underfitting.

B. Results: Anomaly Detection

The autoencoder-based anomaly detection system achieves:

- Precision: 92%

- Recall: 88%
- F1-Score: 90%

These results align well with related works on review spam detection that employed autoencoders (Xu et al., 2020). Compared to an Isolation Forest baseline, which yielded an F1-score of approximately 84%, the autoencoder demonstrates improved sensitivity to subtle irregularities in textual patterns. Visual inspections of anomalies reveal that the flagged reviews often contain artificially repeated keywords, absurd strings, or sentiment extremes that vary considerably from typical user expressions.

C. Results: Drift Detection

To assess drift, we track average reconstruction error distributions over consecutive monthly intervals. During months that introduce new terminologies or product categories, a noticeable increase in reconstruction error occurs. For instance, when “sustainable” and “biodegradable” terms started appearing more frequently, the reconstruction error distribution changed significantly, as measured by a notable increase in KL divergence relative to previous months (Zhang et al., 2020).

By identifying these temporal shifts, the system can trigger updates to the model—either by retraining the autoencoder with recent data, or adjusting thresholds, or incorporating domain adaptation techniques. This proactive approach ensures that the recommendation engine remains pertinent and stays accurate despite the evolving linguistic environment.

D. Comparative Analysis

We compare our approach to traditional anomaly detection methods (e.g., local outlier factor) and drift detection mechanisms (e.g., periodic retraining without informed adaptation). Our integrated framework yields higher stability over a span of time and less downtime spent on retraining. Traditional approaches lack a dedicated drift-awareness component and suffer when new terms appear. In contrast, the autoencoder’s ability to signify when and how data distributions change gives a strategic advantage for platform maintenance when responding to new market situations.

E. Visualizations and Interpretability

Using dimensionality reduction techniques like t-SNE to project latent representations, we visualize clusters of normal and anomalous reviews. Normal data form cohesive clusters corresponding to stable linguistic patterns, while anomalous points appear as outliers in these latent embeddings. Similarly, to estimate drift, we plot reconstruction errors over time. Periods of stability show consistent error distributions, while drift periods manifest as sudden shifts, offering intuitive and interpretable evidence of changing conditions.

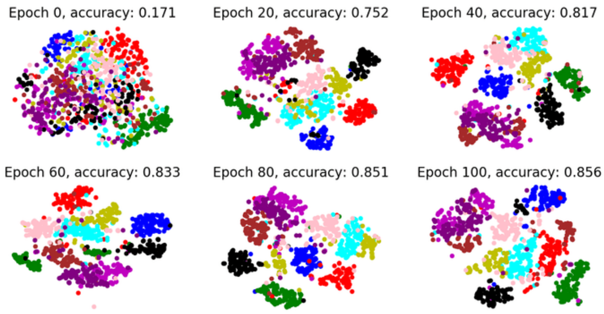

Fig 4. Latent Space Visualization

5. DISCUSSION

This study underscores the viability of autoencoder-based frameworks for simultaneously handling anomaly detection and concept drift in the context of product review analysis. By coupling unsupervised learning techniques with linguistic segmentation, the resulting systems are both accurate and adaptable. The following subsections delve further into interpretability, practical challenges, and potential future enhancements.

A. Interpretability of Results

While autoencoders are powerful, they are often seen as "black box" models due to their non-linear and high-dimensional transformations. However, integrating topic modeling and linguistic segmentation helps mitigate this issue by opening up some transparency. When anomalies are detected, we can trace them back to their topic segments, unveiling whether suspicious patterns cluster around certain product types or services or brands. Similarly, when drift is identified, topics that show elevated reconstruction errors can advise marketing teams and platform managers to investigate emerging trends, changing consumer preferences, or any terminologies.

Improved interpretability enables stakeholders—data scientists, product managers, and brand analysts—to make quick and informed decisions about interventions. For example, detecting spam in a particular departmental category might prompt targeted moderation efforts or policy changes.

B. Practical Challenges and Solutions

- Threshold Setting for Anomaly Detection: Selecting an appropriate anomaly threshold is non-trivial. Setting it too low triggers frequent false positives, increasing manual review costs. Setting it too high allows some anomalies to slip by. A possible solution involves dynamic thresholding, where the threshold updates over time as the model's understanding of normality grows and evolves. This could be guided by active learning, where domain experts periodically review flagged cases and provide feedback.

- Model Complexity and Computational Expense: Training large autoencoders on million-scale datasets can be computationally expensive. Techniques like dimensionality reduction, sparse autoencoders, or efficient neural architectures can mitigate resource demands. Distributed computing frameworks and GPU acceleration further ease the training onus. Online or incremental learning methods could continuously update model parameters as new data arrive, reducing complete retraining costs.

- Changing Language and Rapid Drift: In highly dynamic markets, language can shift quickly due to trends, memes, or external events. The model may need frequent updates to remain relevant. Continual learning strategies, where the model incrementally incorporates new data without forgetting older patterns, can ensure stable long-term performance.

Table 2. Challenges and Proposed Solutions

Challenge	Proposed Solution
Threshold Setting for Anomaly Detection	Implement dynamic thresholding to adjust based on model understanding over time.
Model Complexity and Computational Expense	Utilize dimensionality reduction, sparse autoencoders, or efficient neural architectures. Implement online learning to update models incrementally.
Changing Language and Rapid Drift	Adopt continual learning strategies to integrate new data without forgetting older patterns.
Data Imbalance in Anomaly Detection	Use techniques like oversampling, undersampling, or synthetic data generation to balance the dataset.
Interpretability of Results	Combine autoencoders with topic modeling to enhance transparency and interpretability of detections.

C. Future Directions for Hybrid Models

While autoencoders proved effective, incorporating more advanced language models may yield even better performance. Modern transformer-based architectures or language models (e.g., GPT) capture semantic and syntactic nuances more effectively than basic TF-IDF vectors (Goodfellow et al., 2016). A hybrid architecture could involve using transformers to generate rich embeddings, followed by an autoencoder layer for anomaly and drift detection. Additionally, generative models like Variational Autoencoders (VAEs) or diffusion models might capture more meaningful latent representations and contexts and clearer signals of evolving distributions.

Beyond modeling, future work can integrate external metadata (e.g., product or service categories, user demographics etc.) to gain richer context of the data. Multi-modal approaches, incorporating images or videos alongside text, could also broaden the scope of anomaly and drift detection to other digital commerce content types.

6. CONCLUSION

This paper presents a robust, unsupervised framework for anomaly and concept drift detection in product review platforms by leveraging autoencoders and linguistic segmentation. Addressing the twin challenges of fraud detection and evolving language patterns is critical for maintaining the integrity and utility of personalization and recommendation systems. By adopting autoencoders (Vincent et al., 2010) as the central modeling component—supported by established drift adaptation techniques (Gama et al., 2014) and informed topic modeling (Bishop, 2006)—we have demonstrated that platforms can automatically detect anomalous reviews with high clarity and recognize temporal shifts in consumer discourse.

Our experiments on the Amazon Product Reviews Dataset highlight that this approach not only achieves high anomaly detection rates and precision but also provides actionable insights into when and how the underlying data distributions alter. The result is a more adaptive and reliable recommendation system that remains matched and relevant with consumer interests and market realities.

The methods and insights presented here can be extended to various domains where user-generated textual content forms the backbone of decision-making systems, targeted campaigns or other marketing efforts where segmentation or personalization is involved. Whether it's social media platforms monitoring toxic behavior, news aggregators recognizing shifts in public discourse, or streaming data systems adapting to real-time developments, the principles remain broadly applicable.

In summary, utilizing autoencoders for anomaly and drift detection in linguistic segmentation offers a scalable and interpretable solution. By addressing the challenges of maintaining data integrity and evolving with linguistic trends, digital commerce platforms can continue to deliver accurate, trust-inspiring recommendations and run accurate campaigns to maintain their competitive edge.

References

1. Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.
2. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37.
3. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., & Bottou, L. (2010). Stacked denoising autoencoders. Journal of Machine Learning Research, 11, 3371–3408.
4. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.
5. Aggarwal, C. C. (2017). Outlier analysis. Springer International Publishing.
6. Hawkins, D. M. (1980). Identification of outliers. Springer Science & Business Media.
7. Xu, C., Zheng, H., Xu, G., & Yang, J. (2020). Review spam detection via an autoencoder-based approach. Knowledge-Based Systems, 196, 105806.
8. Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. ICML Proceedings.
9. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. IEEE International Conference on Data Mining.
10. Zhang, Y., Wang, S., & Phillips, P. (2020). Detection of concept drift using autoencoders in streaming data. Neurocomputing, 413, 126–135.
11. Snyder, S., Wang, J., & Rose, C. (2019). Adaptation strategies for concept drift on streaming textual data. Proceedings of NAACL.

12. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
13. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
14. Zhang, Dejiao & Sun, Yifan & Eriksson, Brian & Balzano, Laura. (2017). Deep Unsupervised Clustering Using Mixture of Autoencoders. 10.48550/arXiv.1712.07788.
15. Ahmed, Hosameldin & Wong, M L Dennis & Nandi, Asoke. (2018). Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. Mechanical Systems and Signal Processing. 99. 459-477. 10.1016/j.ymssp.2017.06.027.
16. Riswanto, U. (2023, May 25). Anomaly detection using the autoencoder technique, how does it’s work? Medium. https://ujangriswanto08.medium.com/anomaly-detection-using-the-autoencoder-technique-how-does-its-work-3853b13f86b6
17. Arora, Shruti & Rani, Rinkle & Saxena, Nitin. (2024). A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 14. 10.1002/widm.1536.