https://doi.org/10.37375/bsj.v7i20.3657

A Generalized Extreme Value Naïve Bayes Framework for Classifying Univariate Extreme Value

Zahra Alghanay Mohammed ***Souad A. Abumaryam**

تاريخ الاستلام: 9 / 8 / 2025 اجازة النشر:10/ 10/ 2025 تاريخ النشر:17/ 11/ 2025

Abstract: This paper aims to develop a model that predicts the class of extreme values by utilizing the Naïve Bayes algorithm and employing statistical models from Extreme Value Theory (EVT). It aligns with recent interests in creating innovative algorithms that integrate both machine learning (ML) and EVT. The model is developed using simulated data, and the experimental results demonstrate significant performance, achieving an accuracy, sensitivity, and specificity of 0.98, along with an AUC of 0.97. The MATLAB software package is used to model and implement the proposed algorithm, which establishes a classification rule for determining the class of extreme values.

Keywords: Extreme value theory, Generalized Extreme Value distribution, Naïve Baayes, Machine Learning.

إطار بيز الساذج ذي القيمة المتطرفة المعممة لتصنيف القيمة المتطرفة أحادية المتغير

زهرة الغناي محمد أبودربالة **سعاد أحمد أبومريم**

قسم الرياضيات، جامعة سرت

المستخلص: تهدف هذه الورقة الى بناء نموذج تعلم آلة باستخدام لغة الماتلاب، يتنبأ بفئة القيم القصوى من خلال توظيف نموذج Naïve Bayes واستغلال نتائج مهمة وتوزيعات احتمالية تقدمها نظرية القيمة القصوى. يتبع هذا البحث الاهتمامات الحديثة في بناء نموذج تربط بين مجال تعلم الألة ونظرية القيمة القصوى. أظهرت النتائج التجريبية أداءً متميزاً حيث بلغت دقة النموذج وحساسيته ونوعيته 0.98، كما بلغت مساحة تحت المنحنى 0.97.

الكلمات المفتاحية: نظرية القيمة القصوى، التوزيع المعمم للقيم الوسطى، بايز، تعلم الألة.

INTRODUCTION

The accurate modelling and classification of extreme events such as floods, hurricanes, catastrophic storms, financial crashes, unprecedented temperatures, system failures, and others are of paramount importance across numerous scientific and engineering disciplines. These events, lying in the tail of probability distribution, present unique challenges for conventional statistical and machine learning methods, which typically emphasize central tendencies and often treat extremes as noise. Extreme Value Theory (EVT) provides a robust statistical framework for examining the behaviour of distribution tails. However, its integration into adaptive, probabilistic machine learning systems remains limited.

Conventional machine learning approaches often encounter difficulties in addressing with statistical challenges posed by extreme values, as they tend to focus on data points near the mean and often treat extremes as outlier to be removed or normalized (Silverberg and Verspagen, 2007). However, the integration of Bayesian methods with EVT has emerged

Department of Mathematics, University of Sirte, Sirte, Libya z.aghanay@su.edu.ly

∗ Department of Mathematics, University of Sirte, Sirte, Libya s.ahmed@su.edu.ly

as a promising solution, enhancing adaptability and the representation of uncertainty. In the supervised machine learning context, Emtair in (Emtair, 2015) developed classification models based on the peaks-over-threshold approach from (EVT). This model adapts the traditional Bayesian framework by incorporating mixed Generalized Pareto components, achieving high accuracy on simulated data. However, a comprehensive framework for classifying extreme values using the theoretically grounded block-maxima approach within a probabilistic classifier remains underdeveloped. Additionally, in (Emtair, 2015) focused on determining an optimal boundary for the mixed Generalized Pareto distribution, employing Bayesian criteria as a classification method and utilizing the Newton-Raphson method for implementation.

Extreme Value Theory (EVT) provides a rigorous statistical framework specifically designed for modeling rare, high-impact events. Founded on seminal work by Coles in (2001) and others, EVT offers asymptotic distributions for extreme observations, primarily through the Generalized Extreme Value (GEV) distribution for block maxima. The Fisher-Tippett theorem establishes the GEV as the universal limiting distribution for normalized maxima, regardless of the underlying data-generating process. The Fisher Tippett theorem, a cornerstone of EVT, states that the distribution of normalized block maxima converges to a Generalized Extreme Value (GEV) distribution, regardless of the underlying data's distribution (Ferreira & de Haan, 2006).

This theoretical foundation provides a powerful tool for modelling extremes, yet its application has been largely confined to statistical domains, with limited integration into adaptive machine learning systems (Sabourin, 2021). Sabourin in (2021) advocated for incorporating EVT principles into ML frameworks to enhance robustness in tail behaviour modelling.

Integrating Bayesian methods into EVT helps address some limitations of traditional approaches. EVT has emerged as a critical statistical tool for understanding extreme phenomena. It relies on asymptotic arguments to model observations that are significantly larger or smaller than typical values (Coles, 2001). Two primary approaches within EVT are the Block Maxima (BM) and Peaks Over Threshold (POT) methods. The BM method involves dividing a dataset into non-overlapping blocks and identifying the maximum or minimum value within each block, while the POT method focuses on observations that exceed a predetermined threshold (Ferreira & de Haan, 2006).

The Generalized Extreme Value (GEV) distribution, derived from EVT, encompasses three types of extreme value distributions: Gumbel, Fréchet, and Weibull. The GEV distribution's flexibility allows it to model a wide array of extreme phenomena effectively (Reiss & Thomas, 2007). The Fisher-Tippet theorem reinforces the theoretical foundation of EVT by providing conditions under which the distribution of maxima converges to one of the GEV types as the sample size increases (Ferreira & de Haan, 2014).

The Naïve Bayes algorithm is a probabilistic classifier based on Bayes theorem, making it suitable for situations where class distributions can be modeled parametrically or non-parametrically. The algorithm assumes that the features are conditionally independent given the class label, simplifying the computation of class probabilities (Murphy, 2012). This characteristic makes Naïve Bayes particularly effective for classifying extreme events, as it can integrate the GEV distribution for modeling class densities.

This paper is focused on the block-maxima approach as a method for extracting maxima. A Generalized Extreme Value (GEV) distribution is assumed for the extracted maxima, concentrating on one-dimensional feature space. Experiments on simulated data are carried out to provide insights into the mechanisms of the approach and to evaluate the model performance in classification tasks.

1. Theoretical Foundation

In this section we shall introduce some probability distributions that figure prominently in extreme value theory and in applications.

.1. Extreme Value Theory 1

Extreme Value Theory (EVT) is a statistical and theoretical framework that creates modeling tools and statistical models to describe the behavior of extreme and rare events, including the minimum and maximum levels of various processes. This theory has become an important statistical tool in several fields where extreme events occur regularly. EVT introduces a statistical model based on asymptotic arguments, focusing specifically on extreme observations with unusually large values (Coles, 2001; Sabourin, 2021).

1.1.1. Extreme Value Distribution

A random variable X has an extreme value (EV) distribution if and only if its probability distribution function is given by (Reiss, 2007):

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right) \qquad 1+\gamma x > 0$$

Here, γ is known as the shape parameter. In particular, we have:

$$G_\gamma(x) \to G_0(x) \quad as \quad \gamma \to 0.$$

Applying the well-known formula,

$$\Longrightarrow \quad G_0(x) = \exp(-e^{-x}), \quad \forall\, x(1+\gamma x)^{1/\gamma} \to \exp(x) \quad as \quad \gamma \to 0$$

By noting that the density function is the derivative of the distribution one, one can obtain the corresponding densities for the EV distribution functions:

$$g_\gamma(x) = \begin{cases} G_\gamma(x)(1+\gamma x)^{-\left(1+\frac{1}{\gamma}\right)}, & 1+\gamma x > 0,\ \gamma \neq 0 \\ G_\gamma(x)e^{-x}, & x\ ,\gamma = 0 \end{cases}$$

1.1.2. The Generalized Extreme Value Distribution

The Generalized Extreme Value (GEV) distributions unifies the possible limiting distributions for normalized maxima. A random variable z follows a GEV distribution if its cumulative distribution function (CDF) is expressed as follows (Reiss, 2007):

$$G_{\gamma,\mu,\sigma}(z) = \exp\left(-\left(1+\gamma\left(\frac{z-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right), \quad 1+\gamma\left(\frac{z-\mu}{\sigma}\right) > 0$$

- $\gamma \in \mathbb{R}$, is the shape parameter, governing the tail
- $\mu \in \mathbb{R}$, it the location parameter, determining the center of the distribution.
- $\sigma > 0$ is the scale parameter, controlling the spread of the distribution.

The parameter μ determines the position of the distribution on the x-axis, while σ determines its spread. The GEV distribution unifies three extreme value families:

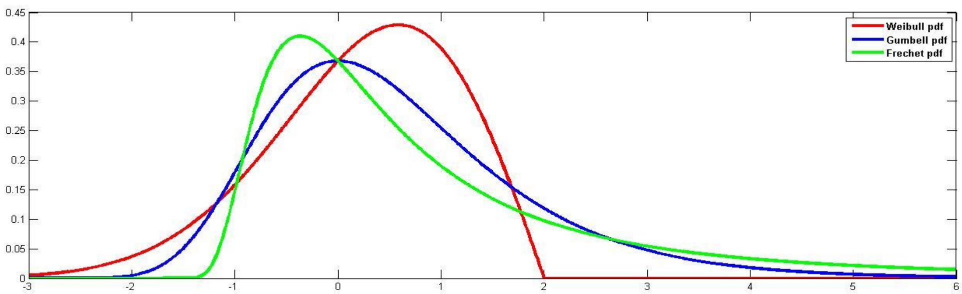
- Gumbel $(\gamma = 0)$: Light-tailed distributions.
- Fréchet $(\gamma > 0)$: Heavy-tailed distributions.
- Gumbel $(\gamma < 0)$: Short-tailed distributions with a finite upper endpoint.

We have

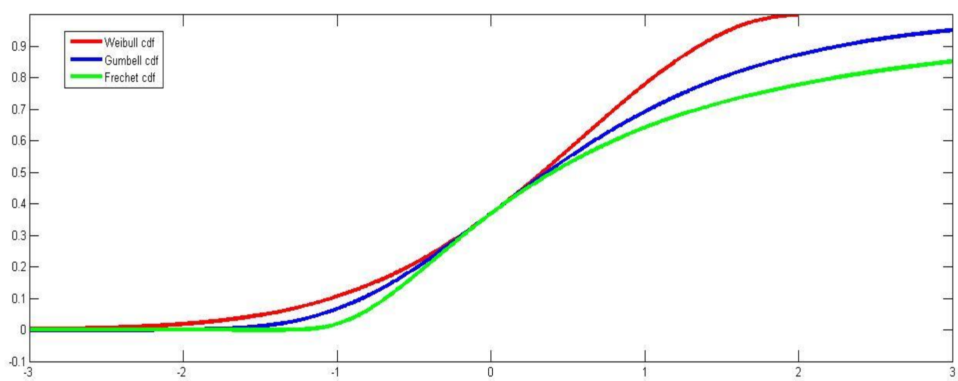
, as $\gamma \to 0 G_{\gamma,\mu,\sigma} \to G_{0,\mu,\sigma}$

The corresponding probability density function is derived as the derivative of the CDF, $G_{0,\mu,\sigma}$ representing generalized Gumbell distribution. The generalized densities are,

$$g_{\gamma,\mu,\sigma}(z) = \begin{cases} \frac{1}{\sigma} e^{-\left(\frac{z-\mu}{\sigma}\right)} \exp\left(-e^{-\left(\frac{z-\mu}{\sigma}\right)}\right) & ,\gamma = 0 ,\forall x \\ \frac{1}{\sigma}\left(1+\gamma\left(\frac{z-\mu}{\sigma}\right)\right)^{-\left(\frac{1}{\gamma}+1\right)} \exp\left(-\left(1+\gamma\left(\frac{z-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right) & ,\gamma \neq 0 \end{cases}$$



Figure(1). *GEV density functions*



Figure(2). *GEV Cumulative Distribution Functions*

1.2. Block-maxima Modelling

The Block-Maxima (BM) approach is one of the two primary method in EVT for extracting extreme values from datasets (the other being the Peaks Over Threshold (POT)). In POT, an observation is deemed extreme if it exceeds a specified threshold, typically modeled with the generalized pareto distribution. The BM method involves dividing the observation period into m nonoverlapping blocks of equal size. The maximum or minimum observation value from each block is then identified as an extreme observation, which is then modeled using the GEV distribution.

Formally, given a sequence of independent observation $X_1, X_2, \ldots$ represent the random sample with distribution function F. the block maxima for k blocks of size m are defined as

$$M_i = \max\{X_j\} \quad , \quad (i-1)m < j \leq im, \quad i = 1, 2 \ldots, k \,,$$

$$m = 1, 2$$

The total number of observations is given by $n = m \times k$. Under certain conditions, these observations can be approximated by an (EV) distribution, G_γ for some real γ. Statistical methods of inference are then applied to analyze these extreme value distributions. A practical advantage of the BM approach is that, it often relies solely on block maxima, such as yearly maxima from extensive historical records. Moreover, this method can be easier to implement, as block periods naturally arise in many contexts, particularly when data confirm to a block structure, such as annual temperature or financial data (Coles, 2001; Ferreira & de Haan, 2014). *Figure 1 illustrates 100 observations divided into 10 blocks, with the red dots indicating the extreme events (block maxima).*

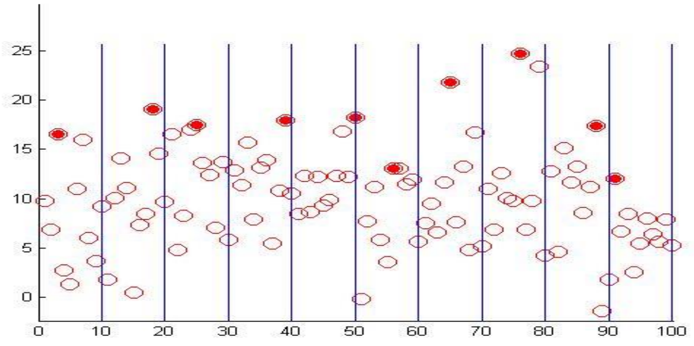


Figure (3). *Extraction of extreme events in the BM approach*

Figure (3) shows 100 observations divided into 10 blocks and from each block; the red dot represents the extreme event (block maxima).

1.3. Fisher Tippet Theorem

The Fisher Tippett Theorem provides the fundamental theoretical justification for modeling block maxima using extreme value distributions. Consider the sample maximum

$$M_n = \max\{X_j\}$$

from a sequence of independent and identically distributed random variables with common distribution F. The exact distribution of M_n can be obtained from the underlying distribution:

$$\Pr\{M_n \leq z\} = \Pr\{X_1 \leq z, \ldots, X_n \leq z\} = \prod_{i=1}^{n} Pr\{X_i \leq z\} = \{F(z)\}^n$$

However, this representation presents two practical challenges. First, the underlying distribution F is typically unknown in real world applications. Second, as $n \to \infty$, the distribution degenerate:

$$\lim_{n\to\infty}\{F(z)\}^n = \begin{cases}0, & z < z+\\ 1, & x \geq z+\end{cases}$$

Where $z+= sup\{x: F(x) < 1\}$ denotes the upper endpoint of F , promoting the approximate models for F^n. This indicate that the distribution of M_n degenerates to a point mass at z+. To obtain a non-degenerate limiting distribution, we consider normalized maxima:

$$M_n^* = \frac{M_n - b_n}{a_n}$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$. Appropriate choices of $\{a_n\}$ and $\{b_n\}$ prevents the limiting distribution to degenerate at one point and hence, the difficulties that arise with the variable M_n are eliminated. The entire range of possible limit distributions for M_n^* is given by the following theorem (Coles, 2001).

The Fisher-Tippet theorem states that if (X_n) is a sequence of independent and identically distributed random variables, and there exist normalizing constants , $a_n > 0$, b_n ∈ R and non–degenerate distribution function G such that:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z) \quad as\, n \to \infty$$

Then G must be a generalized extreme value distribution,

$$G_{\gamma,\mu,\sigma}(z) = \exp\left(-\left(1+\gamma\left(\frac{z-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right), \quad 1+\gamma\left(\frac{z-\mu}{\sigma}\right) > 0$$

This remarkable theorem establishes that the normalized maximum of the random variable M_n converges in distribution to a limiting form that must belong to one of three types of extreme value distribution. This has profound implication for statistical modelling: regardless of the underlying population distribution F, the asymptotic behavior of sample maxima is universally characterized by the GEV distribution family. The three types Gumbel, Fréchet, and Weibull encompass all possible limiting forms for normalized maxima, making the GEV distribution the fundamental modelling framework for extreme value analysis (Coles, 2001), see also (Ktakauer, 2024).

2. Machine Learning Foundation

The exponential growth of data availability has intensified the demand for computational methods that can transform raw information into actionable insight. Machine Learning (ML) addresses this need by developing algorithms capable of extracting meaningful pattern from data to predict future outcomes and support decision making under uncertainty. This interdisciplinary field leverages statistical theory and computational tools to build models that learn from historical observation, enabling informed predictions and data driven actions.

Machine learning employs parametrized models whose structures are determined through optimization of objective functions. These mathematical formulations quantify model performance, guiding the selection of optimal parameters from training data using computational implementations. The resulting optimized model serve as powerful tools for prediction, automation, and decision support across diverse applications.

.1 Theoretical Basis of machine learning2

Machine learning tests on statistical and probabilistic foundation for building mathematical models that extract information about data generating processes from historical observations. From a statistical perspective, training data constitute a sample drawn from an underlying population distribution. The fundamental task of machine learning algorithms is therefore to make inferences about population characteristics from this finite sample, with computer science providing the computational infrastructure to implement statistical inference techniques and manage large scale datasets (Alpaydin, 2010).

.2 Supervised Machine Learning2

Supervised learning represents one of the primary paradigms in machine learning (other primary is unsupervised learning), characterized by learning input-output mapping from labelled training data. Given a datasets D= $\{(x_i, y_i), i = 1,2, \dots, n\}$ of input-output pairs, the goal is to learn a function $f(x, \theta) = y^*$ that accurately predicts outputs for new inputs by optimizing parameters θ according to a performance criterion. Input x_i may represent diverse data types, including numerical feature vectors or complex structured data such as images, text, or graphs. Outputs y_i typically take categorical values from a finite set $y_i \in \{\omega_1, \dots, \omega_c\}$ for classification model. Binary classification represents the special case where c= 2, with models seeking decision boundaries that optimally separate classes to enable accurate prediction (Alpaydin, 2010).

.3 Naïve Bayes classifier2

The Naïve Bayes classifier is a probabilistic classification method based on Bayes theorem, with the fundamental assumption of conditional independent between features given the class label. For binary classification, the algorithm models the training set as realizations of independent and identically distributed random variables with prior probability distribution p(y;θ) and class conditional densities p(x|$y = \omega_j; \Theta_j$); j=1,2 . The class prior follows a Bernoulli distribution y$\sim Ber(\theta)$:

$$p(y;\theta) = \begin{cases} \theta \;, & y = \omega_1 \\ 1-\theta, & y = \omega_2 \end{cases}$$

.3.1 Training stage 2

The training process of the Naïve Bayes classifier involves estimating the class conditional density function p(x|$y = \omega_j; \Theta_j$); j=1,2 and the class prior distribution $p(y)$. Two primary approaches are available for this estimation:

parametric Approach model: This method assumes a specific distributional form for the class conditional densities and employs statistical estimation techniques to make inferences about the parameters Θ_i;i=1,2 and θ from the training data. The fundamental assumption is that all examples within each class are generated from the same underlying distribution. The complete data likelihood function is given by:

$$L(\Theta_1,\Theta_2,\theta) = \prod_{i=1}^{n} p(x_i,y_i) = \prod_{i=1}^{n} p(x_i|y_i)p(y_i)$$

Which factorizes into class specific component:

$$L(\Theta_1,\Theta_2,\theta) = \{\prod_{i=1}^{n_1} p(x_i|y_i = \omega_1;\Theta_1)\, p(y_i = \omega_1;\theta)\} \times \{\prod_{i=1}^{n_2} p(x_i|y_i = \omega_2;\Theta_2)\, p(y_i = \omega_2;\theta)\}$$

In practice, we maximize the log likelihood for computational stability:

$$l(\Theta_1,\Theta_2,\theta) = \sum_{i=1}^{n} \{\log p(x_i|y_i) + \log p(y_i)\}$$
$$= \sum_{i=1}^{n_1} \{\log p(x_i|y_i = \omega_1;\Theta_1) + \log p(y_i = \omega_1;\theta)\}$$
$$+ \sum_{i=1}^{n_2} \{\log p(x_i|y_i = \omega_2;\Theta_2) + \log p(y_i = \omega_2;\theta)\}$$

Where n_1 and n_2 represent the number of training examples in classes ω_1 and ω_2 respectively.

Class priors $p(y)$ are conveniently estimated from class frequencies in the training data.

Nonparametric Approach model: When distributional assumptions are untenable, nonparametric methods such as histograms or kernel density estimation provide flexible alternative for modelling class conditional densities without presupposing specific forms. The parametric approach offers significant advantages by reducing the classification problem to estimating a small number of parameters through established statistical Figures 12-14 illustrate the training process, showing the original data methods. distribution, fitted class conditional densities using Gaussian model, and estimated class priors (Alpaydin, 2010).

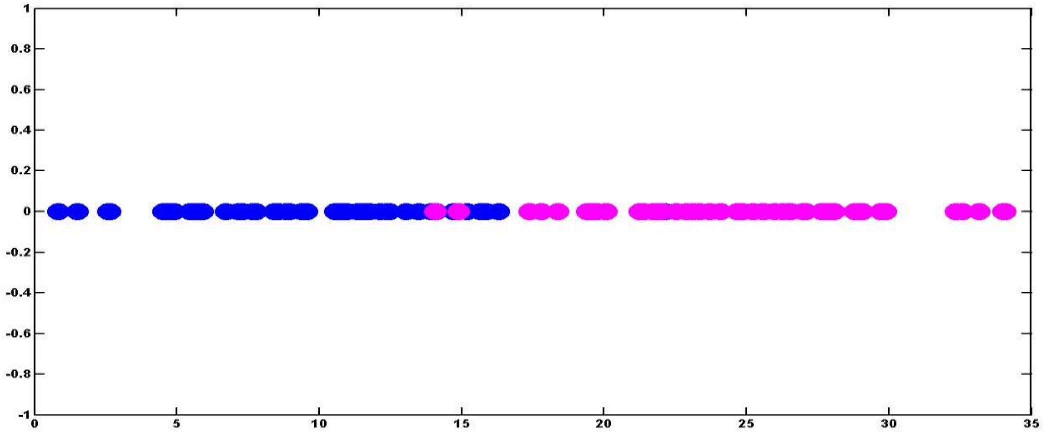


Figure (4). *Representation of Training Set*

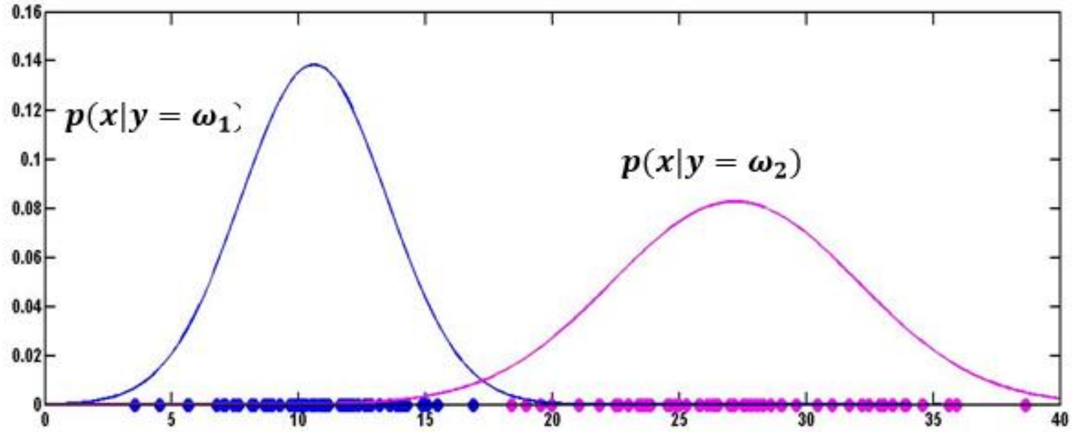


Figure (5). *Modelling the class densities with normal distribution*

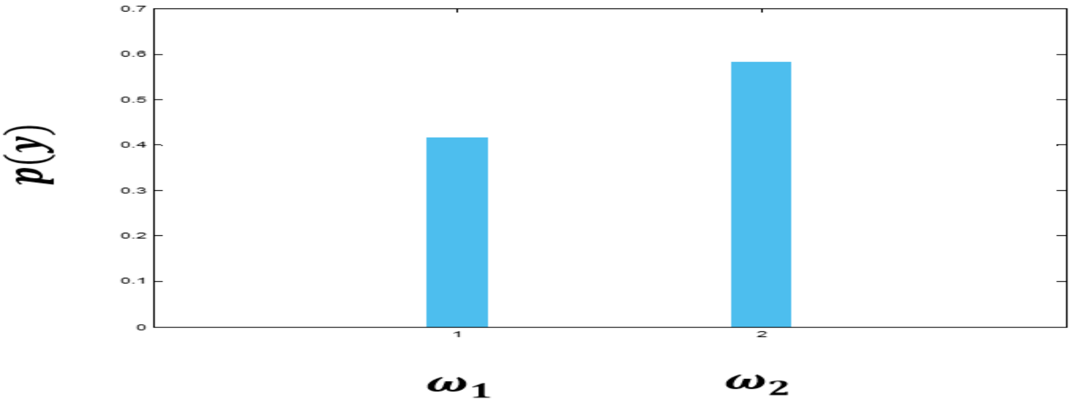


Figure (6). *The Estimated Class Prior*

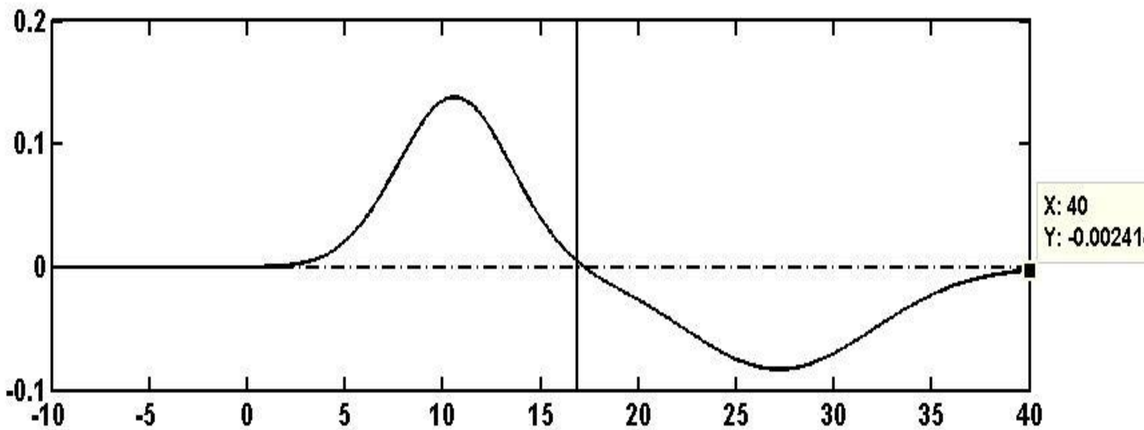


Figure (7). *The Curve of the Class Posteriors Difference*

.3.2 Decision stage2

Once the class conditional densities and prior are estimated, classification of new observations proceeds through Bayesian inference. For a test instance x, we compute posterior probabilities:

$$P(y = \omega_i|x) = \frac{p(x|y = \omega_i)p(y = \omega_i)}{p(x)} \quad ; i = 1,2$$

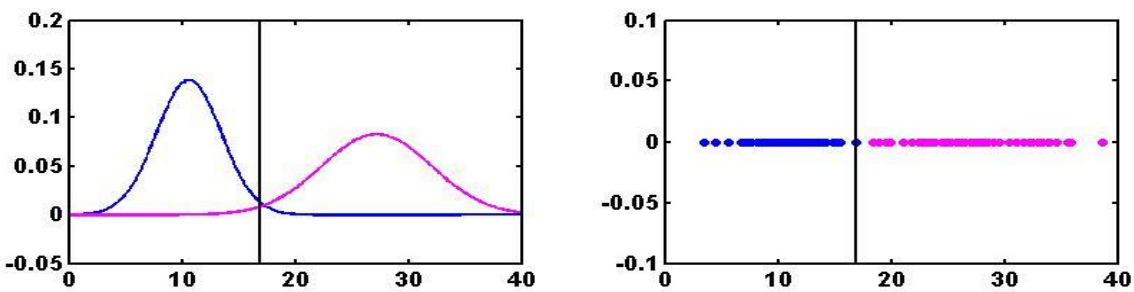
Where the evidence term $p(x) = \sum_{i=1}^{2} p(x|y = \omega_i)p(y = \omega_i)$ ensures proper normalization. The optimal classification rule assigns x to the class with maximum posterior probability:

$$y = \arg\max_{\omega_i} P(y = \omega_i|x)$$

This is equivalent to assigning x to class ω_1 if:

$$p(x|y = \omega_1)p(y = \omega_1) - p(x|y = \omega_2)p(y = \omega_2) > 0$$

Figure 7 and Figure 8 illustrate the difference in class posteriors and the resulting decision boundary that optimally separates the class.



Figure(8). *The Classification Model (decision rule)*

2.3.3 Model Evaluation and Rejection Option

Rejection action:- Classification uncertainty arises when posterior probabilities P($y = \omega_j$|x) are nearly equal across classes, or when the maximum posterior probability, remains relatively small. To address this uncertainty, a rejection can be implemented by introducing a classification threshold θ. Observation x are rejected when the maxima posterior probability satisfies $P(y = \omega_j|x) < \theta$, ensuring that only confident predictions are accepted while ambiguous cases are flagged for further review (Alpaydin, 2010).

Confusion matrix Analysis: Classifier performance is quantitatively evaluated using a confusion matrix constructed from a held-out test set, typically obtained through bootstrap resampling from the original dataset. This matrix systematically compares true class labels against predicted classifications, enumerating true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Table (1). This table of errors is called the confusion matrix.

Table (1). ***Confusion Matrix***

True class ∵predicted class	w_1	w_2
ω_1	TPR	FNR
ω_2	FPR	TNR

From this framework, essential performance metrices can be computed:

True Positive Rate (TPR) (sensitivity) =TP/(TP+FN)

True Negative Rate (TNR) (specificity) =TN/(TN+FP)

False Positive Rate (FPR) (sensitivity) = FP/(TN+FP)

False Negative Rate (RNR) (specificity) = FN/(TP+FN)

These metrices provide a comprehensive assessment of classifier performance across different error types and operational requirement (Murphy, 2012).

3. Methodology

This section outlines the systematic procedure for developing and evaluating the GEV Naïve Bayes classifier for univariate extreme values. The framework encompasses data preparation, maxima extraction via Block Maxima, model specification, parameter estimation, and model evaluation, providing a replicable pipeline for extreme value classification.

3.1 Maxima Extraction: - For a binary classification with classes $\omega_1$and ω_2, we begin a labelled random sample $D = \{(z_i, y_i): y_i \in \{\omega_1, \omega_2\}\}_{i=1}^{n}$ drawn from the population $\chi \times \{\omega_1, \omega_2\}$, where $z_i \in \mathbb{R}$. The class prior is defined as a Bernoulli distribution:

$$\mathrm{p}(\mathrm{y};\theta) = \begin{cases} \theta \;, & y = \omega_1 \\ 1-\theta, & y = \omega_2 \end{cases} \qquad 1$$

Where, the partition:

$$\{(z_i, y_i): z_i \in \mathbb{R}, y_i \in \{\omega_1, \omega_2\}\}_{i=1}^{n} = \bigcup_{j=1,2} \left\{\left(z_{i,\omega_j}, \omega_j\right)\right\}_{i=1}^{nj}$$

where $n_1 + n_2 = n$ and n_j the number of examples in class $\omega_j; j = 1,2$.

To extract extrema values, the Block Maxima method is applied separately to each class. For a class , ω_j containing n_j observations, the data is partitioned into m_j non-overlapping blocks of size s_j, where ($m_j \times s_j = n_j$). The maxima are defined as:

$$x_{i,\omega_j} = \max\left\{z_{k,\omega_j}\right\} \quad ; (\mathrm{i}-1)s_j < \mathrm{k} \le i s_j, \;\; \mathrm{i} = 1,2, \dots, t_j, \mathrm{j} = 1,2$$

These extracted maxima are treated as independent and identically distributed realizations following a GEV distribution, justified by the Fisher Tippett theorem (Coles, 2001). The class conditional density for a maximum x given ω_j is:

$$p(x|\omega_j;\gamma_j,\mu_j,\sigma_j)=\frac{1}{\sigma_j}\left(1+\gamma_j\left(\frac{x-\mu_j}{\sigma_j}\right)\right)^{-\left(\frac{1}{\gamma_j}+1\right)}\exp\left(-\left(+\gamma_j\left(\frac{x-\mu_j}{\sigma_j}\right)\right)^{-\frac{1}{\gamma_j}}\right)$$

(Ferreira & De Haan, 2014).، Where, $\gamma_j\left(\frac{x-\mu_j}{\sigma_j}\right)>0$

The parameters shape γ_j, location μ_j and scale σ_j satisfy $-\infty<\gamma_j<\infty$, $-\infty<\mu_j<\infty$ and $\sigma_j>0$

3.2 Classification Methodology: - The Naïve Bayes classifier is employed, which assigns a new observation x (an extracted maximum) to the class with the highest posterior probability. According to Bayes theorem, the posterior probability for class ω_j is given by:

$$P(y=\omega_j|x)=\frac{p(x|\omega_j;\gamma_j,\mu_j,\sigma_j)p(y=\omega_j)}{\sum_{j=1}^{2}p(x|\omega_j;\gamma_j,\mu_j,\sigma_j)p(y=\omega_j)}\quad ;j=1,2$$

The classification rules is defined as:

$$,\quad f:x\to\{\omega_1,\omega_2\}f(x)=\underset{\omega_i}{\operatorname{argmax}}P(y=\omega_i|x)$$

based on the extracted maxima

$$\bigcup_{j=1,2}\left\{(x_{i,\omega_j},\omega_j)\right\}_{i=1}^{nj}$$

This is equivalent to assigning x to a class ω_1 if the discriminant function $B(x)>0$, and to a class ω_2 otherwise, where

$$B(x)=p(x|\omega_1;\gamma_1,\mu_1,\sigma_1)p(y=\omega_1)-p(x|\omega_2;\gamma_2,\mu_2,\sigma_2)p(y=\omega_2)>0$$

This rule is assumed to minimize the conditional risk

$$R(f(x)=\omega_i|x)=\sum_{j=1}^{2}\lambda_{ij}P(y=\omega_i|x);i=1,2$$

Subject to the zero-one loss function,

$$\lambda_{ij}=\begin{cases}0 & ;\ i=j\\ 1 & ;i\neq j\end{cases}$$

Subsequently, the classification rule minimizes the overall risk, (Duda et al, 2001)

$$R=\int R(f(x)|x)\,p(x)dx$$

3.3 Parameters estimation:- The parameters of the GEV Naïve- Bayes model γ_j,μ_j,σ_j; j=1,2 and θ are estimated through the maximum likelihood function, which seeks to find the parameter values that maximize the joint probability of observing the extracted block maxima and their associated class labels. The likelihood function for the observed data is given by:

$$L(\gamma_1, \mu_1, \sigma_2, \gamma_2, \mu_2, \sigma_2, \theta) = \{\prod_{j=1}^{2} \prod_{i=1}^{n_j} p\left(x_{i,\omega_j} \middle| \omega_{j_i}; \gamma_j, \mu_j, \sigma_j \right) p(y = \omega_j; \theta)$$

This is solved numerically to obtain optimal parameter estimates. We seek the parameter values that maximize this likelihood function (Alpaydin, 2004). The maximum likelihood criteria lead to the following system of equations:

$$\sum_{i=1}^{n_j} \{\frac{1}{\gamma_j^2} \ln(y_{i,j}) \left[1 - (y_{i,j})^{-\frac{1}{\gamma_j}}\right] + \left(\frac{x_{i,\omega_j} - \mu_j}{\sigma_j}\right) \left[\frac{1}{\gamma_j} (y_{i,j})^{-(1+\frac{1}{\gamma_j})} - \frac{1}{(\gamma_j + 1) y_{i,j}}\right]\} = 0$$

$$\frac{1}{\sigma_j} \sum_{i=1}^{n_j} [\frac{\gamma_j}{(\gamma_j + 1) y_{i,j}} - (y_{i,j})^{-(1+\frac{1}{\gamma_j})}] = 0$$

$$-\frac{n_j}{\sigma_j} + \frac{1}{\sigma_j} \sum_{i=1}^{n_j} [\frac{\gamma_j}{(\gamma_j + 1) y_{i,j}} - (y_{i,j})^{-(1+\frac{1}{\gamma_j})}] \left(\frac{x_{i,\omega_j} - \mu_j}{\sigma_j}\right) = 0$$

$$, \frac{n_1}{\theta} - \frac{n_2}{1-\theta} = 0$$

re

$$y_{i,j} = 1 + \frac{\gamma_j}{\sigma_j} (x_{i,\omega_j} - \mu_j)$$

Solving this system yields the estimated values of the parameters (Martine & Stedinger, 2005).

3.4 Parametric Model diagnostic tools- : The critical assumption that the block maxima follow GEV distribution is validated using two diagnostic tools:

The Quantile- Quantile(Q-Q) Plots**:** The fitted GEV density is overlaid on a non-parametric kernel density estimate of the data. Close agreement between the two curves supports the GEV assumption.

The Quantile- Quantile(Q-Q) Plots: The empirical quantiles of the extracted maxima are plotted against the theoretical quantile of the fitted GEV distribution. Linearity indicates a good fit.

3.5 Classification boundary detection : The classification boundary (denoted by $x^{[cb]}$) is defined as the solution to equation $B(x) = 0$ for the classification rule $f: \chi \to \{\omega_1, \omega_2\}$, is defined as:

$$x^{[cb]} = \{x \in \chi: B(x) = 0\}$$

Substituting GEV density functions and prior probabilities, this becomes

$$B(x) = \frac{1}{\sigma_1}\left(1+\gamma_1\left(\frac{x-\mu_1}{\sigma_1}\right)\right)^{-\left(\frac{1}{\gamma_1}+1\right)} \exp\left(-\left(1+\gamma_1\left(\frac{x-\mu_1}{\sigma_1}\right)\right)^{-\frac{1}{\gamma_1}}\right) - \frac{1}{\sigma_2}\left(1+\gamma_2\left(\frac{x-\mu_2}{\sigma_2}\right)\right)^{-\left(\frac{1}{\gamma_2}+1\right)} \exp\left(-\left(1+\left(\frac{x-\mu_2}{\sigma_2}\right)\right)^{-\frac{1}{\gamma_2}}\right)(1-\theta) = 0$$

Due to the nonlinear nature of the equation, the bisection method is employed over the interval $[a_1\ a_2]$ that contains $x^{[cb]}$ with a tolerance $\varepsilon = 0.00001$ to numerically approximate the solution. The process involves replacing $a_k, k = 1,2$ by $l = \frac{a_1 + a_2}{2}$ whenever $B(a_k) \times B(l) > 0$, until $a_2 - a_1 < \varepsilon$, for predetermined error ε.

3.6 Model Evaluation Metrics To assess the performance of the proposed model across various threshold values, we evaluate True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Key metrics will include Accuracy, Specificity, Sensitivity, F1-score, ROC curve, PR curve, and AUC score. The AUC score evaluates the model's precision quality independent of the chosen classification threshold.

4. Experimental results

4.1 Simulation Study

To evaluate the performance of the GEV Naïve Bayes framework, we conducted a simulation study. Synthetic datasets were generated sampling 5000 independent observations for each class from two normal distribution: $\omega_1 \sim N(30,5)$ and $\omega_2 \sim N(40,10)$. The block maxima approach was applied with a block size $s_j = 10;\ \ j = 1,2$ for both classes. The specific statistical property of the generated date is summarized in Table 1. Our experiments reveal significant insights into the classification performance, with various diagnostic plots affirming the effectiveness of our GEV-based classification model. Figure 1 shows different plots of the extracted maxima, with block size

T**able2**. Statistical properties of synthetic data

	mean	**Standard deviation**
ω_1	30	5
ω_2	40	10

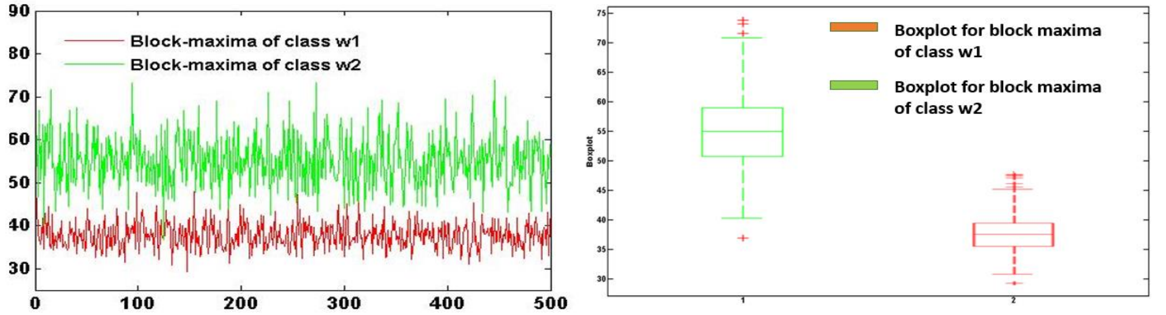


Figure (9): Extracted maxima (left). Boxplot of the extracted maxima in each group (right).

4.2 Parameter Estimation

The GEV parameters were estimated using 70% of the extracted maxima (350 per class), with the remaining 30% for testing.

Table 3 shows the GEV parameter estimates alongside with 95% confidence interval. The estimated values are the results of the maximum likelihood criteria for 70% of the extracted maxima from each class. Various diagnostic plots for assessing the accuracy of the GEV model fitted to the extracted maxima in each class are shown in Figure 18. In each class the Q-Q plot is near linear giving strong evidence for GEV as underlying distribution for the extracted maxima the corresponding GEV density estimate of each class seems consistent with the data as can be seen from the comparison with the kernel density function.

Table 3. GEV parameter estimate.

	γ	μ	σ
ω_1 **95% CI**	-0.1686 (-0.2258.-0.1114)	36.6348 (36.3118.0.9571)	2.9181 (2.7039,3.1494)
ω_2 **95% CI**	-0.1984 (-0.2558, -0.1411)	52.4329 (51.7768,53.089)	5.8945 (5.4584,6.3653)

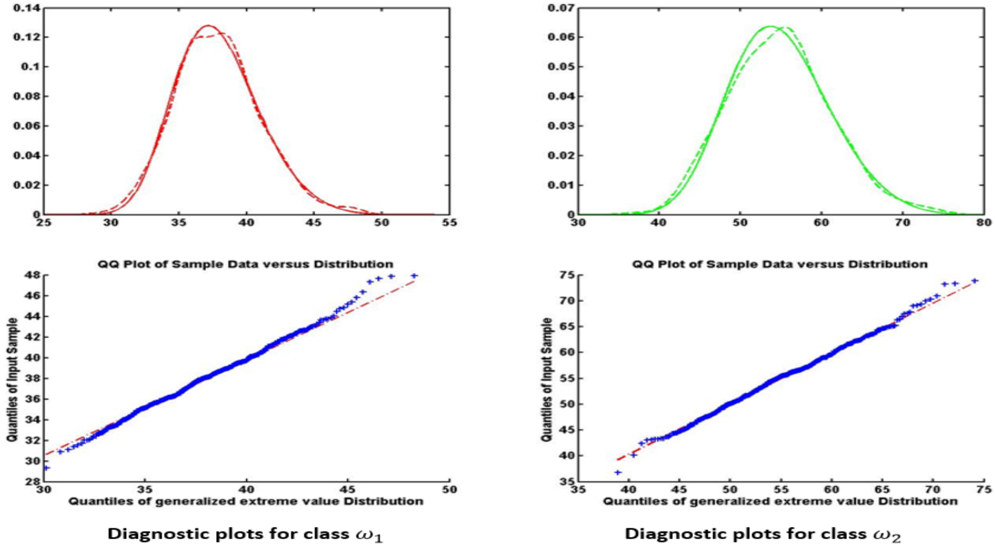


Figure 10. Diagnostic plots

The negative shape parameters for both classes indicate a Weibull type distribution with finite upper endpoints, consistent with the theoretical behaviour of block maxima derived from normal distributions.

4.3 Model Diagnostics

Diagnostic plots assessed the adequacy of the GEV distribution for modelling the extracted maxima. Q-Q plots showed strong linear alignment along the diagonal, while comparative density plots showed close agreement between the fitted GEV distributions and non-parametric kernel.

5.4 Classification Performance

By iterating through the bisection procedure, the classification boundary is obtained to be $x^{[cb]} = 44.6698$. The convergence behaviour of the iteration method is as given in Table3.

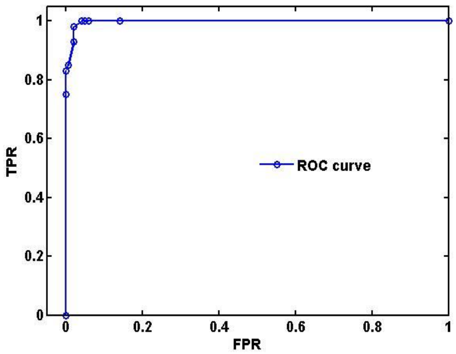
Table 4. Convergence of Bisection Procedure.

iteration	l	$a_2 - a_1$	$B(l)$
1	50.0000	20.000	-0.0526
5	40.3750	1.250	0.0034
10	44.6680	0.391	0.0000
15	44.6698	0.0012	0.0000
18	44.6698	0.0001	0.0000

Table 5 shows the performance of the classification model on 30% of the extracted maxima based on different metrics.

Table (5) Model Performance for different values of threshold.

k	TP	FP	TN	FN		FPR-	Specify -	Sensitivity (TPR)	accuracy	F1-score	precision
0.0	150 -	150 -	0	0		1.00	0.00	1.00	0.5	0.66	0.5
0.1 -	150	21	129-	0		0.14	0.86	1.00	0.93	0.93	0.87
0.2	150	10	140	0		0.06	0.93	1.00	0.96	0.96	0.93
0.3	150	8	142	0		0.05	0.94	1.00	0.97	0.97	0.94
0.4	150	6	144	0		0.04	0.96	1.00	0.98	0.98	0.96
0.5	148	3	147	2		0.02	0.98	0.98	0.98	0.98	0.98
0.6	140	3	147	10		0.02	0.98	0.93	0.95	0.95	0.97
0.7	128	1	149	22		0.01	0.99	0.85	0.92	0.91	0.99
0.8	125	0	150	25		0.00	1.00	0.83	0.91	0.90	1.00
0.9	113	0	150	37		0.00	1.00	0.75	0.87	0.85	1.00
1.0	0	0	150	15 0-		0.00	1.00	0.00	0.50	0.00	1.00



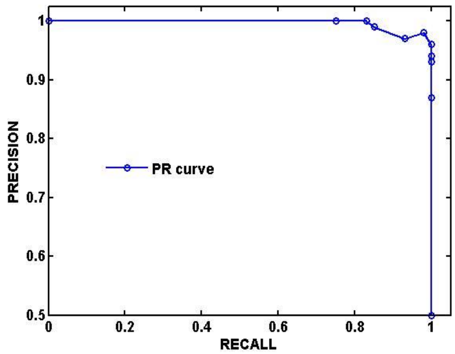


Figure 11. Receiver Operation Curve and Precision-Recall Curve

The performance is measured for different values of threshold k. The threshold values k=0.1-0.4 yield the highest sensitivity rates, while the specifity and precision are the highest for k=0.8, 0.9. The accuracy is happened to be the highest for threshold values k=0.4, 0.5. The f1-score is the highest at k=0.4, 0.5. At the threshold k=0.5 the f1-score, sensitivity, spcifity and accuracy is 0.98. The different values for FPR and TPR (sensitivity) for different values of the threshold k are summarized in the Receiver Operation Curve (ROC). The corresponding values for AUC are 0.97. The area under the Precision-Recall (PR) curve is 0.94.

The high AUC of 0.97 demonstrate the model's strong ability to discriminate between extremes of the two classes across all possible classification thresholds.

5. Discussion

The experimental results demonstrated the effectiveness of GEV-Naïve Bayes framework for the classification of extreme values. The model achieved exceptional performance, with accuracy, specificity, sensitivity, and F1 score all reaching 0.98 at the optimal decision threshold. This indicates that the method effectively distinguishes between extreme values from different classes.

The negative shape parameter estimates for both classes align with theoretical expectations, as block maxima from normal distribution should converge to a Gumbel distribution (a special case of the GEV with =0) in the limit. The finite sample estimates close to zero with negative values are consistent with this expectation. The computed classification boundary effectively separates the distributions of block maxima from the two classes, with class ω_2 (mean 40, SD 10) producing generally larger extremes than class ω_1 (mean 30, SD 5), as expected.

6. Conclusion

In conclusion, it has successfully developed a GEV Naïve Bayes framework for classifying univariate extreme value. This approach merges the principles of EVT with probabilities classifier, ensuring that the distinctive nature of extreme observations is directly incorporated into the model. Using synthetic data, we employed the block maxima method and split the data for training and evaluation. The GEV parameters for each class were estimated via maximum likelihood, and the validity of the model was confirmed through diagnostic plots. The approach leverages the theoretical foundation of Fisher-Tippett theorem and implements a practical classification system through maximum likelihood estimation and numerical optimization for decision boundary detection. The final classification rule based on posterior probabilities derived from Bayes rule, demonstrated remarkable accuracy across all performance metrics.

The primary for future work is to extend this framework to multivariate case, leveraging multivariate EVT to classify extreme events characterized by multiple interacting variables.

(**Acknowledgment**:- We are deeply grateful to Prof Emtair. M. Muharhar for his guidance, insights, and support, which have profoundly influenced our lives)

References

1. Alpaydın, E. (2010). *Introduction to machine learning* (2nd). Massachusetts: MIT Press.
2. Castillo-Mateo, J., Asín, J., Cebrián, A. C., Mateo-Lázaro, J., & Abaurrea, J. (2023). Bayesian variable selection in generalized extreme value regression: Modeling annual maximum temperature. MDPI.
3. Coles, S. (2001). *An introduction to statistical modeling of extreme values* (3rd ed.). London: Springer.
4. Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: John Wiley & Sons.
5. Emtair, M. (2015). On a supervised classifier that can learn better from mixed peaks-over-threshold. *International Journal of Research Studies*.
6. Emtair, M. (2016). Numerical solution to the problem of optimal boundary of generalized Pareto mixtures. *International Journal of Applied Science and Engineering Technology*.
7. Ferreira, A., & de Haan, L. (2006). Extreme Value Theory: An Introduction. New York: Springer.
8. Ferreira, A., & de Haan, L. (2014). On the block maxima method in extreme value theory: PMM estimators. *The Annals of Statistics*.
9. Huet, N., & Prosdocimi, I. (2025). Robust and efficient estimation for the generalized extreme-value distribution with application to flood frequency analysis in the UK.
10. Ktakauer, N. (2024). Extended blended generalized extreme value distribution. Journal of Statistical Distributions and Applications, 12(3), 123-145.
11. Murphy, K. (2012). Machine Learning: A Probabilistic Perspective. Massachusetts: MIT Press.
12. Martin, E., & Stedinger, J. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*.
13. Sabourin, A. (2021). Extreme value theory and machine learning: A year of research at Telecom Paris. *Telecom Paris*.
14. Reiss, R., & Thomas, M. (2007). Statistical Analysis of Extreme Values. Switzerland: Birkhäuser Verlag.