NSS

Neutrosophic Sets and Systems, Vol. 82, 2025

University of New Mexico

Uncertainty-Driven Entrepreneurial Career Education Decision-Making in College Incubators: A Neutrosophic Vague N-Soft Set Approach

Bingquan Yin, Yali Hou*, Rui Zhao, Xiangge Liu

Qinhuangdao Vocational and Technical College, Qinhuangdao, 066100, Hebei, China

*Corresponding author, E-mail: houyali@qvc.edu.cn

(Bingquan Yin: E-mail: Yinbingquan7667@163.com; Rui Zhao: E-mail: vigi0815@qvc.edu.cn; Xiangge Liu: E-mail: Liuxiangge@qvc.edu.cn)

Abstract: Entrepreneurship education has been demonstrating an important role in shaping the career trajectories of university students. Nevertheless, deciding the entrepreneurial potential leftovers as a complicated procedure because of inherent uncertainty in student characteristics. In this study, we propose a Neutrosophic Vague N-Soft Set-based Gradient Boosting (NVNSS-GB) approach to offer an uncertainty-driven decision about the likeliness of students to become entrepreneurs. Our approach introduces novel application NVNSS theory to transform student attributes into truth, falsity, and indeterminacy membership intervals to provide effective handling of vagueness and hesitancy in their responses. Then, we introduce an N-Soft Set classification method to assign a multi-level graded evaluation to university students. Next, we fed the generated enriched feature representation to GB to decide on if a student is likely to be an entrepreneur. Comprehensive analyses on a real case study is conducted using sensitivity analysis, feature importance ranking, ROC-AUC evaluation, and complexity analysis, demonstrated the efficiency, effectiveness, and interpretability of NVNSS-GB. Our findings implied that NVNSS-GB has promising power in improving decision-making within entrepreneurship education, offering an insightful tool for career guidance that can allow universities to design targeted interventions to foster entrepreneurial skills among students.

Keywords: Neutrosophic Sets (NSs), Soft Sets, Neutrosophic Vague N-Soft Sets, Multi-valued Logic, Entrepreneurship Education, College Incubators.

1. Introduction

Entrepreneurship has become key driver of economic growth and innovation, particularly in the context of local college incubators that support student-led startups [1]. These incubators provide a ecosystem for aspiring entrepreneurs to experiment with business ideas, develop prototypes, and seek funding [2]. Nevertheless, the entrepreneurial process is characteristically uncertain, since decision-makers have to navigate unpredictable market conditions, evolving consumer demands, and financial constraints. The previous decision-making (DM) frameworks often fall short in addressing the complexity and vagueness inherent in early-stage entrepreneurship. To better support student entrepreneurs, there is a pressing need for advanced computational models that could handle indeterminate, and conflicting information in DM processes [3], [4].

One of the primary challenges in entrepreneurial ecosystems is the inability to effectively quantify uncertainty in DM. Many business models rely in conventional probability-based approaches that assume complete or well-defined information [5]. Nevertheless, real-world entrepreneurial decisions are inclined with multiple conflicting factors, like market dynamics, funding availability, and competitor strategies—many of which are uncertain and partially acknowledged [6]. Also, decision-makers in college incubators, who are first-time entrepreneurs, might struggle with limited access to reliable data, subjective risk perceptions, and cognitive biases. These challenges necessitate a more flexible, intelligent DM framework that could process vague data more excellently.

Neutrosophic logic, by Florentin Smarandache, extended the classical logic systems with inclusion of concept of truth, falsity, and indeterminacy [7], [8]. Unlike earlier systems relied on fixed degrees of truth and falsity, neutrosophic logic allowed for the representation of uncertainty, contradiction, and incomplete information in DM. This makes it particularly suitable for complex domains like entrepreneurship, where real-world decisions often contain ambiguous or conflicting elements.

In the context of entrepreneurship, neutrosophic logic [7] can model uncertain investment decisions, ambiguous customer responses, and unpredictable market shifts, making it an ideal tool for risk assessment and strategic planning in college incubators. However, while neutrosophic logic provides a powerful foundation, it needs to be integrated with structured DM models to be effectively applied in business settings. Building upon the principles of neutrosophic logic, Neutrosophic Vague N-Soft Sets (NVNSS) [9] is a promising solution for making decision analysis based on the evaluation of alternatives under vague, uncertain, and incomplete conditions [9].

Motivated by that, this study aims to explore novel application of NVNS in addressing uncertainty-driven decision-making challenges faced by student entrepreneurs in college incubators. Specifically, we propose a Neutrosophic Vague N-Soft Set-based Gradient Boosting (NVNSS-GB) approach to offer an uncertainty-driven decision about the likeliness of students to become entrepreneurs. Our approach introduces novel application NVNSS theory to transform student attributes into truth, falsity, and indeterminacy membership intervals to provide effective handling of vagueness and hesitancy in their responses. Then, we introduce an N-Soft Set classification method to assign a multi-level graded evaluation to university students. We introduce a real case study on student entrepreneurs in college incubators, and provide in depth exploration of the data to get useful infights about features and factors affecting the entrepreneurial DM**.** Next, using this case study, we drive novel application of our hybrid NVNSS approach to demonstrate the effectiveness in handling uncertainty through real-world scenarios of college incubators. Proof of concept experimental comparisons NVNSS-GB approach decision models with traditional against set of competing methods highlight the advantages of our solutions and their promise to entrepreneurial settings.

The left part of this paper is structured as follows. Section 2 discusses basic concepts and related literature. Section 3 presents the proposed NVNSS-based DM framework and its mathematical implementation of

NVNSS in entrepreneurial DM. Section 4 discussed the experimental results and comparative analysis. Section 5 presents key findings.

2. Background and Basics

This section provides the fundamental concepts and theoretical background necessary to understand the proposed methodology. This introducing mathematical foundations of Neutrosophic Sets, Vague Sets, Soft Sets, and N-Soft Sets, leading to the formulation of Neutrosophic Vague N-Soft Sets (NVNSS).

Definition 1 ([8]**):** Given universe set $U = \{x_1, x_2, \cdots, x_n\}$, A neutrosophic set, A_{NS}, can be defined with of triplet membership functions namely Truth: $T_{A_N}(x)$, Indeterminacy: $I_{A_N}(x)$, and Falsity: $F_{A_N}(x)$.

$$A_{NS} = \left\{\langle x, \left(T_{A_N}(x), I_{A_N}(x), F_{A_N}(x)\right)\rangle : x \in U \right\} \quad (1)$$

such that

$$^{-}0 \le \sup T_{A_N}(x) + \sup I_{A_N}(x) + \sup F_{A_N}(x) \le 3^{+}.$$

Each membership function is subset of $]0^{-}, 1^{+}[$, but, for real-world application, the following interval [0,1] is used instead.

Definition 2 ([10]**) :** Given universe set $U = \{x_1, x_2, \cdots, x_n\}$, A vague set, A_V, is defined with two membership functions namely one for truth, $T_{VS}(x)$, and other for falseness $F_{VS}(x)$.

$$A_{VS} = \{\langle x, \left(T_{VS}(x), F_{VS}(x)\right)\rangle : x \in U \} \quad (2)$$

Definition 3: Given two vague sets $A = \{x_i, [t_A(x_i), 1 - f_A(x_i)] \mid x_i \in U\}$ and $B = \{x_i, [t_B(x_i), 1 - f_B(x_i)] \mid x_i \in U\}$, then, the following relations can be computed

The union between A and B is defined as follows

$$A \cup B = \{\left(x_i, \left[\max\left(t_A(x_i), t_B(x_i)\right), \max\left(1 - f_A(x_i), 1 - f_B(x_i)\right)\right] \mid x_i \in U \right)\}, \quad (3)$$

The intersection between A and B is defined as follows

$$A \cap B = \{\left(x_i, \left[\min\left(t_A(x_i), t_B(x_i)\right), \min\left(1 - f_A(x_i), 1 - f_B(x_i)\right)\right] \mid x_i \in U \right)\}, \quad (4)$$

The complement of A is defined as follows:

$$A^c = \{(x_i, [f_A(x_i), 1 - t_A(x_i)] \mid x_i \in U)\}. \quad (5)$$

A is subset of if the following condition apply:

$$A \subseteq B \text{ iff } t_A(x_i) \le t_B(x_i), \text{ and } 1 - f_A(x_i) \le 1 - f_B(x_i) \quad (6)$$

Definition 4 ([9], [11]**):** Given universe set $U = \{x_1, x_2, \cdots, x_n\}$, A neutrosophic vague set, A_{NV}, is defined with three membership functions valued as follows:

$$A_{NV} = \{\langle x, (T_{NV}(x), I_{NV}(x), F_{NV}(x))\rangle : x \in U\}$$
$$= \left\{\langle x, \begin{pmatrix} [T_{NV}(x)^-, T_{NV}(x)^+], [I_{NV}(x)^-, I_{NV}(x)^+], \\ [F_{NV}(x)^-, F_{NV}(x)^+] \end{pmatrix}\rangle : x \in U\right\}$$
$$= \left\{\langle x, \begin{pmatrix} [T_{NV}(x)^-, 1 - F_{NV}(x)^-], [I_{NV}(x)^-, I_{NV}(x)^+], \\ [F_{NV}(x)^-, 1 - T_{NV}(x)^-] \end{pmatrix}\rangle : x \in U\right\} \quad (7)$$

Such that

$$^-0 \le T_{A_{NV}}^- + I_{A_{NV}}^- + F_{A_{NV}}^- \le 2^+$$

Definition 5: Given two neutrosophic vague set $A_{NV} = \{\langle x, ([T_{A_{NV}}(x)^-, T_{A_{NV}}(x)^+], [I_{A_{NV}}(x)^-, I_{A_{NV}}(x)^+], [F_{A_{NV}}(x)^-, F_{A_{NV}}(x)^+])\rangle : x \in U\}$ and $B_{NV} = \{\langle x, ([T_{B_{NV}}(x)^-, T_{B_{NV}}(x)^+], [I_{B_{NV}}(x)^-, I_{B_{NV}}(x)^+], [F_{B_{NV}}(x)^-, F_{B_{NV}}(x)^+])\rangle : x \in U\}$, then, the following relations can be computed

The union between A_{NV} and B_{NV} is defined as follows

$$A_{NV} \cup B_{NV} = \left\{\langle x, \begin{pmatrix} [\min(T_{A_{NV}}(x)^-, T_{B_{NV}}(x)^-), \min(T_{A_{NV}}(x)^+, T_{B_{NV}}(x)^+)], \\ [\max(I_{A_{NV}}(x)^-, I_{B_{NV}}(x)^-), \max(I_{A_{NV}}(x)^+, I_{B_{NV}}(x)^+)], \\ [\max(F_{A_{NV}}(x)^-, F_{B_{NV}}(x)^-), \max(F_{A_{NV}}(x)^+, F_{B_{NV}}(x)^+)] \end{pmatrix}\rangle : x \in U\right\}, \quad (8)$$

The intersection between A and B is defined as follows

$$A_{NV} \cup B_{NV} = \left\{\langle x, \begin{pmatrix} [\max(T_{A_{NV}}(x)^-, T_{B_{NV}}(x)^-), \max(T_{A_{NV}}(x)^+, T_{B_{NV}}(x)^+)], \\ [\min(I_{A_{NV}}(x)^-, I_{B_{NV}}(x)^-), \min(I_{A_{NV}}(x)^+, I_{B_{NV}}(x)^+)], \\ [\min(F_{A_{NV}}(x)^-, F_{B_{NV}}(x)^-), \min(F_{A_{NV}}(x)^+, F_{B_{NV}}(x)^+)] \end{pmatrix}\rangle : x \in U\right\}, \quad (9)$$

The complement of A_{NV} is defined as follows:

$$A_{NV}{}^c = \{\langle x, ([1 - T_{A_{NV}}(x)^+, 1 - T_{A_{NV}}(x)^-], [1 - I_{A_{NV}}(x)^+, 1 - I_{A_{NV}}(x)^-], [1 - F_{A_{NV}}(x)^+, 1 - F_{A_{NV}}(x)^-])\rangle : x \in U\} \quad (10)$$

A is subset of if the following condition apply:

$$A \subseteq B_{NV} \text{ iff } t_{A_{NV}}(x_i) \le t_{B_{NV}}(x_i), \text{and } 1 - f_{A_{NV}}(x_i) \le 1 - f_{B_{NV}}(x_i) \quad (11)$$

Definition 6 ([12]): Given universe set $U = \{x_1, x_2, \cdots, x_n\}$, and P be a set of parameters that describe properties, attributes, or criteria relevant to elements of U, then a soft set (F, S) over U is defined as a pair:

$$(F, S): F: S \to P(U)\ \forall\ e \in S \text{ such that } S \subseteq P \quad (12)$$

where S is a finite set of parameters. F is a mapping function.

Example 1: Given $U = \{x_1, x_2, x_3, x_4, x_5\}$ denote a set of students, and $P = \{e_1\text{: Good in math}, e_2\text{: Good in programming}, e_3\text{: Good in communication skills}\}$ is a set of parameters, then, we can drive soft set (S, A) with $A = \{e_1, e_2\}$ could be defined as:

$$F(e_1) = \{x_1, x_3, x_5\}$$
$$F(e_2) = \{x_2, x_4\}$$

This means that the students x_1, x_3, x_5 are good in mathematics, while students x_2, x_4 are good in programming.

Definition 7: Given universe set $U = \{x_1, x_2, \cdots, x_n\}$, and P be a set of parameters that describe attributes, or criteria relevant to elements of U, then a neutrosophic vague soft set (NVSS) (F, T) over U is defined as a pair:

$$\begin{gathered}(F, T),\\ F: T \rightarrow NV(U)\ \forall\ e \in T \text{ such that } T \subseteq P\end{gathered} \tag{13}$$

Definition 8: Given a group of universe sets $U_1, U_2, \ldots, U_n$, and P as a set of parameters, an N-Soft Set, an extension of soft set theory, is defined over this group

$$(F, S): F: S \rightarrow P(U_1) \times P(U_2) \times \ldots \times P(U_n)\ \forall\ e \in S \text{ such that } S \subseteq P \tag{14}$$

where F is a multi-set-valued mapping function

Example 2: Given three different domains of members: students denoted as $U1 = \{x_1, x_2, x_3\}$, mentors denoted as $U_2 = \{y_1, y_2, y_3, y4\}$, and investment firms denoted as$U_3 = \{z_1, z_2, z_3\}$. Also, we have parameters $P = \{e_1$: Entrepreneurial skills, e_2: Technical proficiency}. N-soft set (F, S) over these domains could be defined as:

$$F(e_1) = (\{x_1, x_3\}, \{y_2, y_3\}, \{z_1, z_3\})$$

$$F(e_2) = (\{x_2, x_3\}, \{y_1, y_4\}, \{z_2\})$$

This means that the members that have entrepreneurial skills are students: x_1, x_3, mentors: y_2, y_3, investment firms: z_1, z_3, meanwhile the members that have Technical Proficiency are students x_2, and x_3, mentors: y_1 and y_4, and the investment Firm z_2.

Definition 9 ([9]**):** Given universe set $U = \{x_1, x_2, \cdots, x_n\}$ with set of parameters $P = \{e_1, e_2, \ldots, e_m\}$, subset of parameters $T \subseteq E$ under evaluation, and set of ordered grades $G = \{0,1, \ldots, N-1\}$, then Neutrosophic Vague N-Soft Set (NVNSS), denoted as (NV, K), is defined as a mapping:

$$NV: T \rightarrow \bigcup_{e_j \in T} \mathcal{NV}\left(NV(e_j)\right) \tag{15}$$

where $\mathcal{NV}\left(NV(e_j)\right)$ is a NVSS of parameter $e_j \in T$, and for each object $x_i \in U$ and parameter $e_j \in T$, there is a distinct pair $(x_i, g_{ij}) \in U \times G$. Thus, the definition of NVNSS is expressed as:

$$\mathcal{NV}\left(NV(e_j)\right) = \left\{\left((x_1, g_{1j}), T_{NV}(x_1), I_{NV}(x_1), F_{NV(x_1)}\right), \ldots, \left((x_n, g_{nj}), T_{NV}(x_n), I_{NV}(x_n), F_{NV}(x_n)\right)\right\}$$

Example 2: Consider having three students $U = \{x_1, x_2, x_3\}$, set of attributes E (e_1 =Mathematics Skill, e_2 =Programming Skill, e_3 =Communication Skill), subset of attributes $T = \{e_1, e_2\}$, and set of grades $G = \{0 = \text{Poor}, 1 = \text{Fair}, 2 = \text{Good}, 3 = \text{Excellent}\}$. If we define the NVNSS for mathematics skill criteria as follows.

$$\mathcal{NV}(NV(e_1)) = \left\{\begin{gathered}((x_1, 3), [0.7,0.9], [0.1,0.3], [0.0,0.2]), ((x_2, 2), [0.5,0.8], [0.2,0.4], [0.1,0.3]),\\ ((x3,1), [0.3,0.6], [0.3,0.5], [0.2,0.4])\end{gathered}\right\}$$

This means that student x_1 has excellent mathematics skills, with truth-membership [0.7, 0.9], meaning they are highly skilled. Student x_2 has good mathematics skills, but with some uncertainty (I = [0.2, 0.4]). Student x_3 has fair mathematics skills, with more uncertainty (I = [0.3, 0.5]).

Definition 10 ([9]): Given two NVNSS (NV_1, K_1) and (NV_2, K_2) on the universe discourse $U = \{x_1, x_2, \cdots, x_n\}$, then $(NV_1, K_1) \subseteq (NV_2, K_2)$ if and only if

$$(NV_1, K_1) \subseteq (NV_2, K_2) \text{iff} \begin{array}{c} T^-_{NV_1}(x) \le T^-_{NV_2}(x), T^+_{NV_1}(x) \le T^+_{NV_2}(x); \\ I^-_{NV_1}(x) \ge I^+_{NV_2}(x), I^+_{NV_1}(x) \ge I^-_{NV_2}(x); \\ F^-_{NV_1}(x) \ge F^-_{NV_2}(x), F^+_{NV_1}(x) \ge F^+_{NV_2}(x). \end{array} \tag{16}$$

Definition 11 ([9]): Given two NVNSS (NV_1, K_1) and (NV_2, K_2) on the universe discourse $U = \{x_1, x_2, \cdots, x_n\}$, then (NV_1, K_1) is said to be equal (NV_2, K_2) iff

$$\begin{array}{c} (NV_1, K_1) = (NV_2, K_2) \\ \text{iff} \\ (NV_1, K_1) \subseteq (NV_2, K_2) \ \& \ (NV_2, K_2) \subseteq (NV_1, K_1) \end{array} \tag{17}$$

Definition 12 ([9]): Given an NVNSS (NV_1, K_1) on the universe discourse $U = \{x_1, x_2, \cdots, x_n\}$, then, it can be called neutrosophic absolute vague N-soft set iff.

$$\begin{array}{c} A_{NV}(e_j) = \{< (o_1, N-1), [1,1], [0,0], [0,0] >, \cdots, < (o_n, N-1), [1,1], [0,0], [0,0] >\} \\ \forall e_j \in T \subseteq E. \end{array} \tag{18}$$

Definition 13 ([9]): Given two NVNSS (NV_1, K_1) and (NV_2, K_2) on the universe discourse $U = \{x_1, x_2, \cdots, x_n\}$, then, the following relations can be computed:

The restricted intersection between $(NV_1, K_1) \ \widetilde{\cap}_R \ (NV_2, K_2)$ is defined as follows

$$\begin{array}{c} K_1 \cap_{\mathcal{R}} K_2 = \big(E, T \cap S, \min(N_1, N_2)\big) \forall e_j \in T \cap S \ x_i \in U \\ < (x_i, g_{ij}), \widehat{T}_{NV}(x_i), \widehat{I}_{NV}(x_i), \widehat{F}_{NV}(x_i) > \in \eta_{NV}(e_j) \\ \Leftrightarrow \\ g_{ij} = \min(g^1_{ij}, g^2_{ij}), \\ \widehat{T}_{NV}(x_i) = \left[\min\left(T^-_{NV_1}(x_i), T^-_{NV_2}(x_i)\right), \min\left(T^+_{NV_1}(x_i), T^+_{NV_2}(x_i)\right)\right], \\ \widehat{I}_{NV}(x_i) = \left[\max\left(I^-_{NV_1}(x_i), I^-_{NV_2}(x_i)\right), \max\left(I^+_{NV_1}(x_i), I^+_{NV_2}(x_i)\right)\right], \text{and} \\ \widehat{F}_{NV}(x_i) = \left[\max\left(F^-_{NV_1}(x_i), F^-_{NV_2}(x_i)\right), \max\left(F^+_{NV_1}(x_i), F^+_{NV_2}(x_i)\right)\right]. \end{array} \tag{19}$$

The restricted union $(NV_1, K_1) \cup (NV_2, K_2)$ is defined as follows

$$\begin{array}{c} K_1 \cup_{\mathcal{R}} K_2 = \big(E, T \cap S, \min(N_1, N_2)\big) \forall e_j \in T \cap S \ x_i \in U \\ g_{ij} = \max(g^1_{ij}, g^2_{ij}), \\ \widehat{T}_{NV}(x_i) = [\max(T^-_{NV_1}(x_i), T^-_{NV_2}(x_i)), \max(T^+_{NV_1}(x_i), T^+_{NV_2}(x_i))], \\ \widehat{I}_{NV}(x_i) = [\min(I^-_{NV_1}(x_i), I^-_{NV_2}(x_i)), \min(I^+_{NV_1}(x_i), I^+_{NV_2}(x_i))], \text{and} \\ \widehat{F}_{NV}(x_i) = [\min(F^-_{NV_1}(x_i), F^-_{NV_2}(x_i)), \min(F^+_{NV_1}(x_i), F^+_{NV_2}(x_i))]. \end{array} \tag{20}$$

The complement of (NV_1, K_1) is defined as $(NV_1, K_1)^c \ N_1 \ \widetilde{\cap} \ N_1^c = \phi$, where NV_1^c

$$NV_1^c = \begin{matrix} \hat{T}_{NV_1}^c(x_i) = [1 - T_{NV_1}^+(x_i), 1 - T_{NV_1}^-(x_i)] = [F_{NV_1}^-(x_i), F_{NV_1}^+(x_i)], \\ \hat{I}_{NV_1}^c(x_i) = [1 - I_{NV_1}^+(x_i), 1 - I_{NV_1}^-(x_i)], \\ \hat{F}_{NV_1}^c(x_i) = [1 - F_{NV_1}^+(x_i), 1 - F_{NV_1}^-(x_i)] = [T_{NV_1}^-(x_i), T_{NV_1}^+(x_i)]. \end{matrix} \tag{21}$$

3. Proposed Methods

In this section, we introduce the methodology of the proposed NVNSS-integrated approach for modeling entrepreneurial competency based on uncertainty and vagueness in student characteristics while utilizing ML models for accurate classification.

Given $U = \{x_1, x_2, \dots, x_n\}$ be the set of students, and $E = \{e_1, e_2, \dots, e_m\}$ be the set of evaluation attributes (features). Given a dataset D consisting of $|E| = 16$ features, we perform the following preprocessing steps:

In step 1, we replace missing values with "Unknown", while for numerical features, we impute missing values with median. Then, we apply a function $f: C \to \mathbb{Z}$ where each categorical variable is assigned a unique integer label such that $C = \{c_1, c_2, \dots, c_k\}$ be the set of categorical variables.

In step 2, we use a numerical feature $X = \{x_1, x_2, \dots, x_n\}$, we apply Min-Max normalization:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{22}$$

This encodes all numerical values into interval [0,1].

In step 3, for each numerical feature X, we compute the NS memberships as follows

$$\begin{gathered} T^-(x_i) = x_i \\ T^+(x_i) = 1 - (1 - x_i) \times 0.5 \end{gathered} \tag{23}$$

Similarly, we measure falsity membership to signify the degree to which a student lacks a characteristic:

$$\begin{gathered} F^-(x_i) = 1 - x_i \\ F^+(x_i) = 1 - T^-(x_i) \end{gathered} \tag{24}$$

The indeterminacy membership function models uncertainty due to conflicting evidence:

$$\begin{gathered} I^-(x_i) = 1 - \frac{T^-(x_i) + F^-(x_i)}{2} \\ I^+(x_i) = 1 - \frac{T^+(x_i) + F^+(x_i)}{2} \end{gathered} \tag{25}$$

In step 3, we incorporate N -Soft Set arrangement, in which each feature X is discretized into N stages:

$$G(x_i) = \arg\min_k \left\{ x_i \in \left[\frac{k}{N}, \frac{k+1}{N} \right] \right\},\ k \in \{0,1,2, \dots, N-1\} \tag{26}$$

This way, we assign soft classification level $G(x_i)$ according to N categories.

In step 4, we divide data into features X and target variable y (entrepreneurship potential). The data is split into 80% training and 20% testing:

$$(X_{train}, X_{test}, y_{train}, y_{test}) = \text{split}\ (X, y, \text{test\_size} = 0.2) \tag{27}$$

In step 5, we start building our ML model

$$\hat{y} = f(X_{train}) \tag{28}$$

where f is trained using Gradient Boosting [13]. The main algorithmic steps is as follows

In step 5.1, we begin with an initial constant predictor with loss $L(y_i, F(x))$:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

$$\text{where} \tag{29}$$

$$L(y_i, F(x)) = -\left[y_i \log \sigma\big(F(x)\big) + (1 - y_i) \log\Big(1 - \sigma\big(F(x)\big)\Big)\right]$$

such that $\sigma(F(x)) = \frac{1}{1+e^{-F(x)}}$

In step 5.2, we calculate pseudo residuals, which indicate the direction of improvement

$$r_i^{(m)} = -\frac{\partial L\big(y_i, F(x_i)\big)}{\partial F(x_i)} \rightarrow \; m = 1, \dots, M \tag{30}$$

For log loss, the gradient is:

$$r_i^{(m)} = y_i - \sigma\big(F_{m-1}(x_i)\big)$$

In step 5.3, we fit a weak learner (decision tree) $h_m(x)$ to the residuals:

$$h_m = \arg\min_{h} \sum_{i=1}^{n} \left(r_i^{(m)} - h(x_i)\right)^2 \tag{31}$$

In step 5.4, we compute the optimal shrinkage factor γ_m :

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\big(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\big) \tag{32}$$

For log loss, we do the following computation:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} -\left[y_i \log\, \sigma\big(F_{m-1}(x_i) + \gamma h_m(x_i)\big) + (1 - y_i) \log\, \Big(1 - \sigma\big(F_{m-1}(x_i) + \gamma h_m(x_i)\big)\Big)\right] \tag{33}$$

In step 5.5, we bring up to date the boosted model, while learning rate $\eta (0 < \eta \leq 1)$ is used to slow learning

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

$$F_m(x) = F_{m-1}(x) + \eta \gamma_m h_m(x) \tag{34}$$

We continue repeating steps 5.2 to 5.5 until reaching maximum iterations M or convergence.

In step 6, we evaluate the performance measured using the following metrics:

$$\text{Accuracy} = \frac{\sum \mathbf{1}(y_i = \hat{y}_i)}{|y|}, \text{Precision} = \frac{TP}{TP+FP}, \text{ Recall} = \frac{TP}{TP+FN} \tag{35}$$

4. Case Study & Exploratory Analysis

This study focuses on analyzing entrepreneurial competency among university students in India. The dataset, collected in 2019 for research purposes, serves as the foundation for this case study. The primary objective is to apply our approach to predict whether a student is likely to become an entrepreneur based on their personal traits, experiences, and psychological attributes [14]. The dataset consists of 219 entries and 17 columns (16 input features + 1 target variable). It captures various factors influencing students' entrepreneurial competency, including Age, Gender, Education Sector, City, Personality & Psychological Traits, Health & Mental Well-being, External Influence, Barriers to Entrepreneurship, and Individual Project Experience. Target Variable (y) indicates whether a student is likely to become an entrepreneur (1: Yes, 0: No).

	Age	Perseverance	DesireToTakeInitiative	Competitiveness	SelfReliance	StrongNeedToAchieve	SelfConfidence	GoodPhysicalHealth	y
count	219.0000	219.0000	219.0000	219.0000	219.0000	219.0000	219.0000	219.0000	219.0000
mean	19.7534	3.3516	3.6210	3.5890	3.7215	3.9087	3.5753	3.5616	0.4155
std	1.2898	0.9952	1.1525	1.1109	1.0536	1.0231	1.1201	1.1003	0.4939
min	17.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000
25%	19.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	0.0000
50%	20.0000	3.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	0.0000
75%	20.0000	4.0000	5.0000	4.5000	5.0000	5.0000	4.0000	4.0000	1.0000
max	26.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	1.0000

To better get the relationship between educational background and entrepreneurial inclination, we analyze the distribution of students who are likely to become entrepreneurs versus those who are not across various educational sectors. Figure 1 presenting a bar plot illustrating these distributions, highlighting variations in entrepreneurial interest among students from diverse academic disciplines.

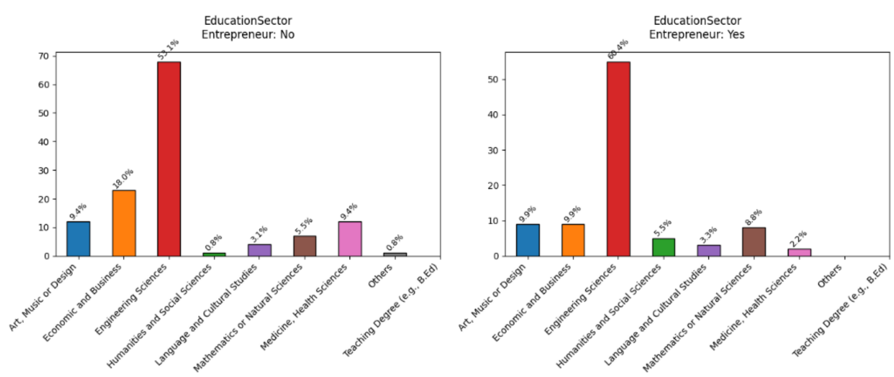


Figure 1. Percentage distribution of entrepreneurial and non-entrepreneurial students based on key personality traits.

To gain deep insight about demographic characteristics of entrepreneurial and non-entrepreneurial students, we conducted a demographic analysis focusing on factors such as gender, age groups, and city of residence. Figure 2 presented pie charts to visualize the proportion of entrepreneurs and non-entrepreneurs within these demographic classes, providing a clearer understanding of how demographic factors influence entrepreneurial inclination.

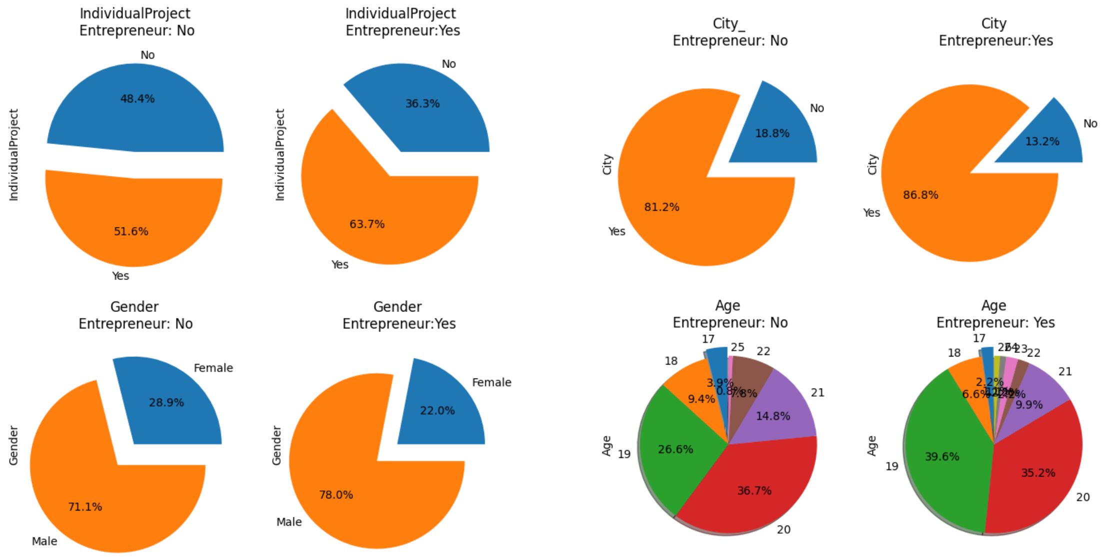


Figure 2. Distribution of key psychological traits among entrepreneurial and non-entrepreneurial students.

Soft skills always play a crucial role in shaping entrepreneurial competency. As a result, we analyze the impact of different soft skills on entrepreneurial potential, we inspected the relationship of entrepreneurial and non-entrepreneurial students w.r.t different soft skills. Figure 3 presented charts that visualize these distributions, providing insights into which soft skills are more prevalent among students likely to turn out to be entrepreneurs.

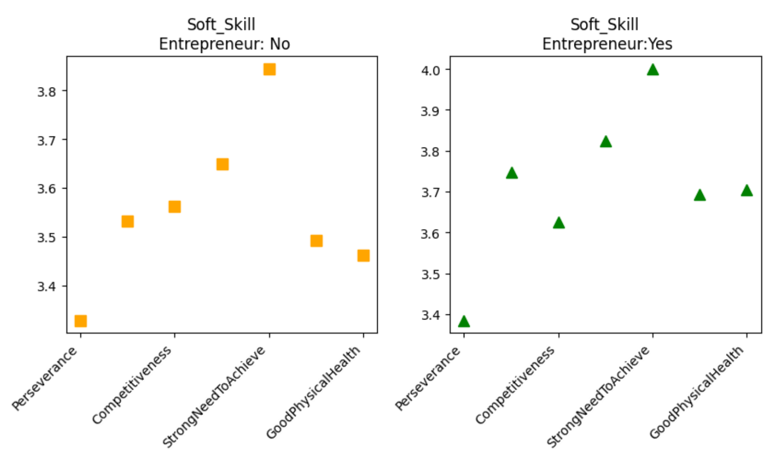


Figure 3. Proportion of entrepreneurial and non-entrepreneurial students based on key soft skills.

Key personality traits always influence an individual's entrepreneurial potential. Hence, we analyzed the distribution of entrepreneurial and non-entrepreneurial students based on various key traits. Figure 4 show these distributions, which highlight the differences in key traits between students who may be entrepreneurs and those who may not.

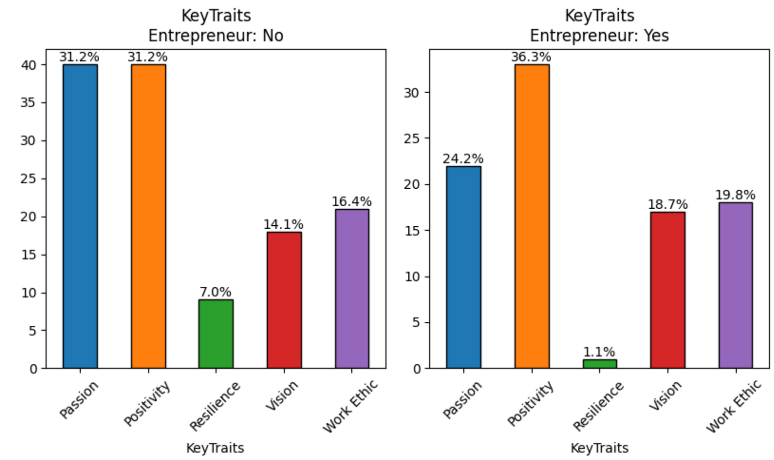
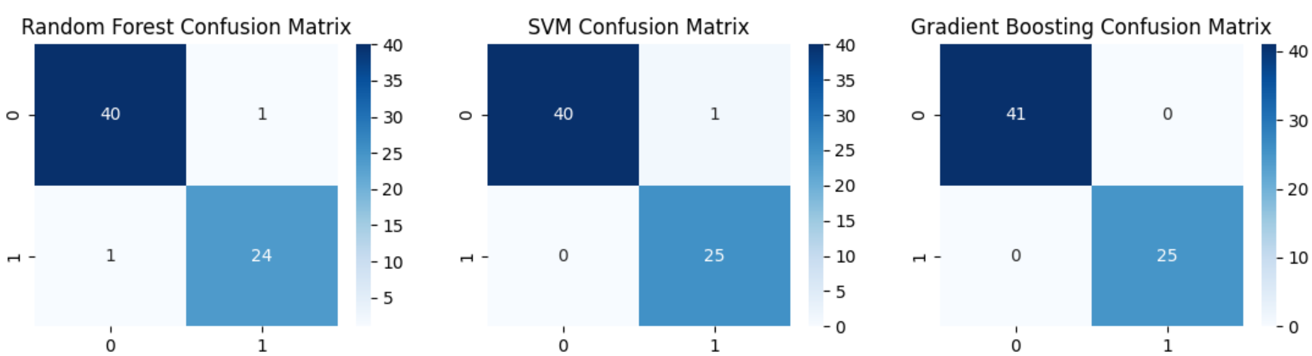


Figure 4. Distribution of key personality traits among entrepreneurial and non-entrepreneurial students.

5. Results & Discussions

In this section we present and discuss different types of analysis to validate the proposed approach. To begin, we compare the performance of NVNSS-GB against other NVNSS-based classifiers in Figure 5 which display their confusion matrices. As shown, the confusion matrix of NVNSS-GB shows a higher number of true positives and true negatives, indicating that it correctly classifies more students with entrepreneurial potential. On the other hand, the alternative NVNSS-based classifiers demonstrated a higher failure rate, (i.e., positive and false negatives). This observed behavior can be attributed to its gradient-based optimization, which iteratively refines weak learners and minimizes classification errors by reducing the impact of indeterminacy and vague memberships in student characteristics. In addition, NVNSS-GB successfully captures nonlinear patterns in the dataset.



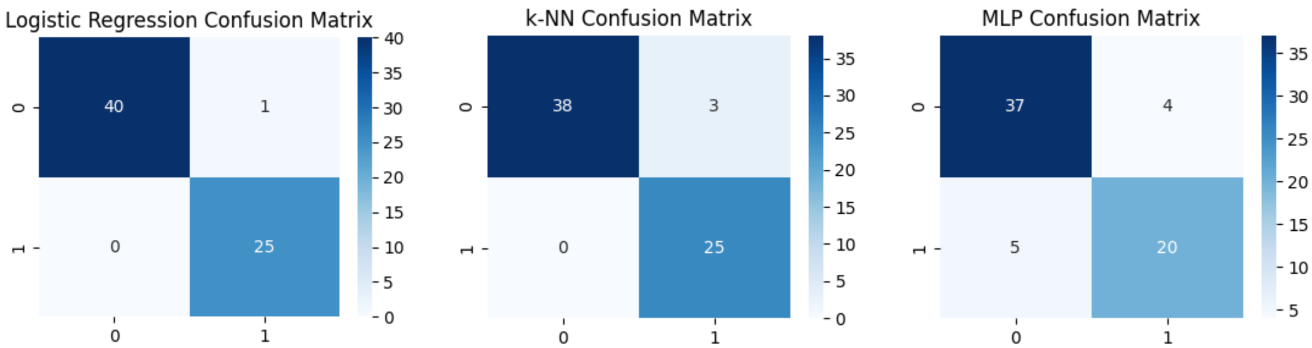


Figure 5. Comparison of Confusion Matrices: NVNSS-GB vs. NVNSS-Classifiers"

In Figure 6, we presented a sensitivity analysis heatmap to illustrate the influence of changing the learning rate and the amount of estimators included. The results indicate that lower learning rates (0.01–0.1) deliver stable performance, while increasing the number of estimators improves accuracy up to a certain threshold, beyond which overfitting occurs.

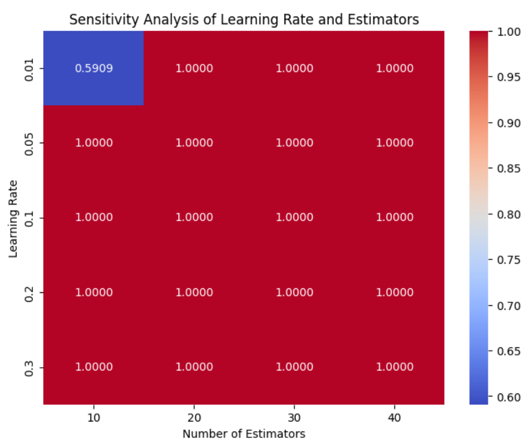


Figure 6. Impact of Learning Rate and Number of Estimators on NVNSS-GB Performance.

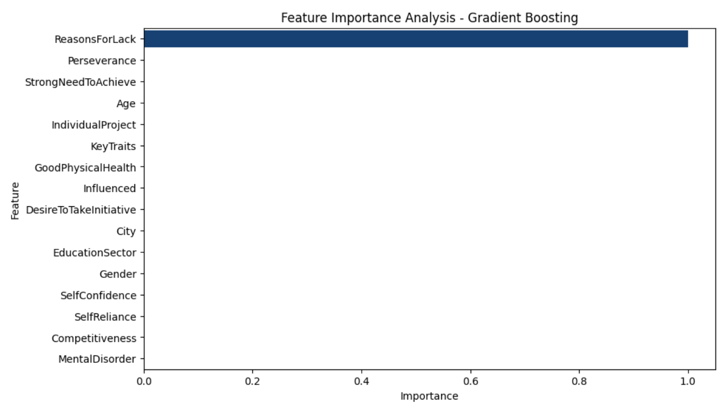


Figure 7. Feature Importance Ranking Using Gradient Boosting and SHAP Values

In addition, we introduce Figure 7 to communicate the feature importance ranking gotten from inner Gradient Boosting importance metrics. Key entrepreneurial traits is indicated as reasons for lack show the highest influence on classification decisions.

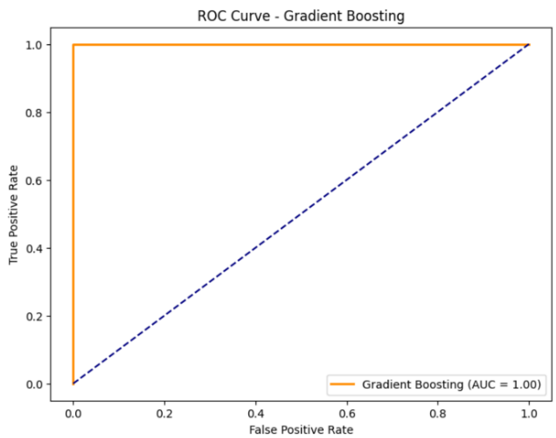


Figure 8. ROC Curve and AUC Score for NVNSS-GB Model.

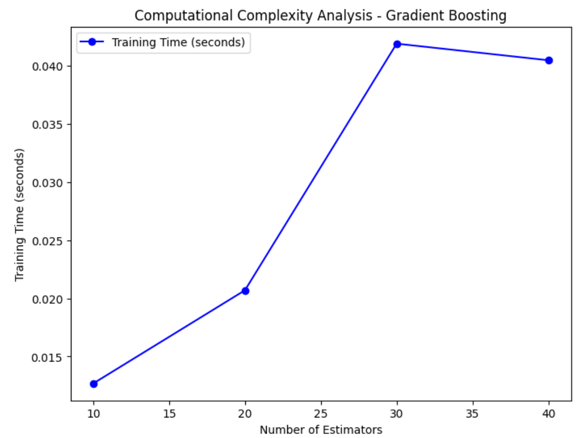


Figure 9. Training Time vs. Number of Estimators in NVNSS-GB.

Besides, Figure 8 presented NVNSS-GB's Receiver Operating Characteristic (ROC), which show that AUC score is 1.00, which indicates the model's strong discrimination ability, effectively distinguishing between students with entrepreneurial potential and those without. In Figure 9, we visualized the relation between training time of NVNSS-GB and number of estimators. As anticipated, increasing the number of estimators causes long training times, with high accuracy improvements.

To investigate instance-level interpretability of the NVNSS-GB, Figure 10 provided SHAP force plot to explain how individual features contribute to a specific prediction, showing whether they push the classification decision towards entrepreneural or non-entrepreneural potential. This effectively demonstrating that ReasonsforLack positively influences predictions.

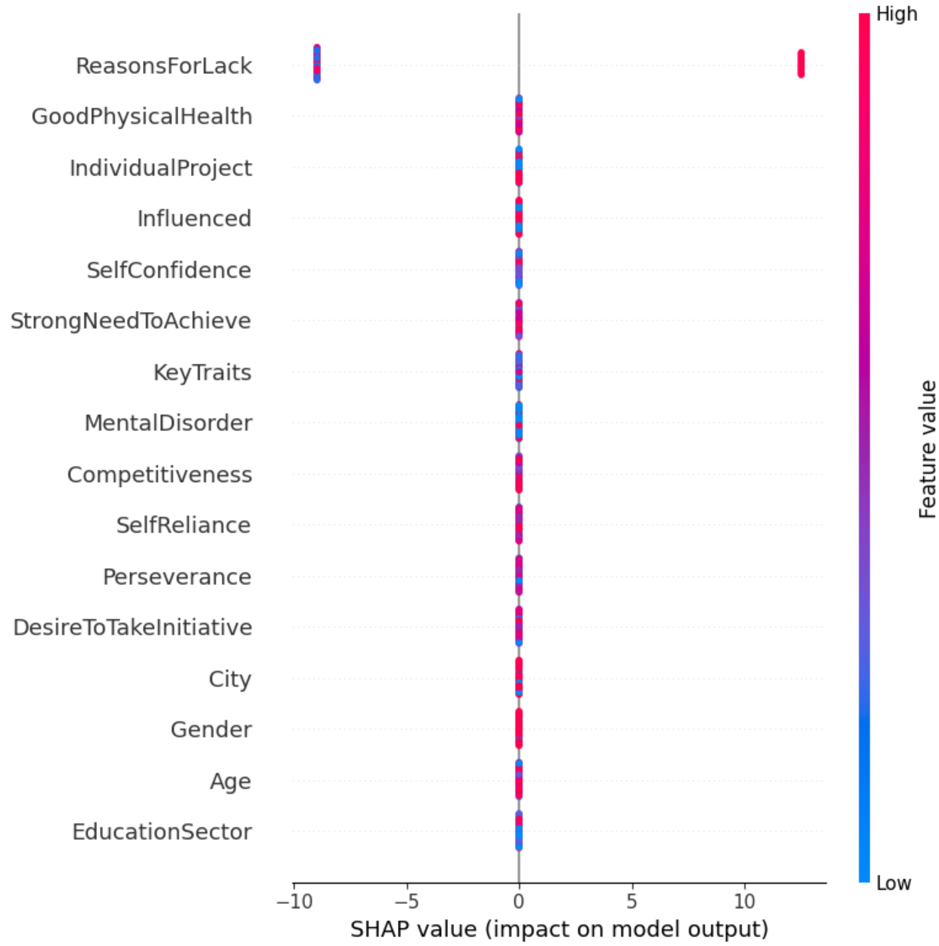


Figure 11. SHAP Force Plot for instance-level Impacts on Entrepreneurial Classification.

6. Conclusion

In this study, we proposed an innovative hybrid approach that integrate Gradient Boosting and NVNSS in a unified framework for enhancing the modeling and optimization of entrepreneurial potential among university students. The NVNSS-based transformation effectively captured uncertainty, vagueness, and indeterminacy in student characteristics, enabling a more robust feature representation. The experimental results on public dataset demonstrate that NVNSS-GB outperforms other NVNSS-based classifiers, as apparent in the confusion matrix. The achieved results imply that the gradient-based optimization can enhances decision boundaries as well as improves the handling of vague membership functions in complex DM scenarios. Future research will emphasize investigating the integration of NVNSS with advanced deep-learning solutions [15], [16].

References

[1] O. Stoica, A. Roman, and V. D. Rusu, "The nexus between entrepreneurship and economic growth: A comparative analysis on groups of countries," *Sustain.*, 2020, doi: 10.3390/su12031186.

[2] N. Bosma, J. Content, M. Sanders, and E. Stam, "Institutions, entrepreneurship, and economic growth in Europe," *Small Bus. Econ.*, 2018, doi: 10.1007/s11187-018-0012-x.

[3] M. Chliova, J. Mair, and A. Vernis, "Persistent Category Ambiguity: The case of social entrepreneurship," *Organ. Stud.*, 2020, doi: 10.1177/0170840620905168.

[4] W. Shi, "Early-stage entrepreneurs' communicative strategies in managing knowledge ambiguity," *J. Small Bus. Manag.*, pp. 1–29, Sep. 2024, doi: 10.1080/00472778.2024.2397411.

[5] Z. Cao and X. Shi, "A systematic literature review of entrepreneurial ecosystems in advanced and emerging economies," *Small Bus. Econ.*, 2021, doi: 10.1007/s11187-020-00326-y.

[6] Z. Xu and S. Dobson, "Challenges of building entrepreneurial ecosystems in peripheral places," *J. Entrep. Public Policy*, 2019, doi: 10.1108/JEPP-03-2019-0023.

[7] F. Smarandache, *A Unifying Field in Logics: Neutrosophic Logic, Neutrosophy, Neutrosophic Set, Neutrosophic Probability*. 1999.

[8] F. Smarandache, *Neutrosophy: neutrosophic probability, set, and logic: analytic synthesis & synthetic analysis*. 1998.

[9] J. Liu, Y. Chen, Z. Chen, and Y. Zhang, "Multi-Attribute Decision Making Method Based on Neutrosophic Vague N-Soft Sets," *Symmetry (Basel).*, 2020, doi: 10.3390/SYM12050853.

[10] W.-L. Gau and D. J. Buehrer, "Vague sets," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, no. 2, pp. 610–614, 1993, doi: 10.1109/21.229476.

[11] A. Al-Quran and N. Hassan, "Possibility Neutrosophic Vague Soft Set for Medical Diagnosis Decision under Uncertainty," *Thai J. Math.*, 2021.

[12] D. Molodtsov, "Soft set theory - First results," *Comput. Math. with Appl.*, 1999, doi: 10.1016/s0898-1221(99)00056-5.

[13] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, 2013, doi: 10.3389/fnbot.2013.00021.

[14] U. Sharma and N. Manchanda, "Predicting and improving entrepreneurial competency in university students using machine learning algorithms," in *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, 2020. doi: 10.1109/Confluence47617.2020.9058292.

[15] W. Abdullah, "Advanced Fruit Quality Assessment using Deep Learning and Transfer Learning Technique," *Sustain. Mach. Intell. J.*, vol. 10, pp. 37–49, Jan. 2025, doi: 10.61356/SMIJ.2025.10450.

[16] W. Ullah, H. Naveed, and S. Ali, "Deep Learning for Precise MRI Segmentation of Lower-Grade Gliomas," *Sustain. Mach. Intell. J.*, vol. 10, pp. 23–36, Jan. 2025, doi: 10.61356/SMIJ.2025.10449.

Received: Oct 11, 2024. Accepted: March 20, 2025