Enhancing Intelligence Source Performance Management through Two-Stage Stochastic Programming and Machine Learning Techniques

Wafula, Lucas Wekesa - Student ID: 190586[a,*]

[a]*Strathmore Institute of Mathematical Sciences*
Strathmore University Ole Sangale Road, Nairobi, 59857-00200, Kenya

Abstract

The effectiveness of intelligence operations is closely tied to the reliability and performance of human intelligence (HUMINT) sources. However, source behavior is often uncertain, deceptive or context-dependent, posing significant challenges for intelligence officers tasked with resource allocation and tasking decisions. This study introduces a hybrid framework that integrates Two-Stage Stochastic Programming (TSSP) with Machine Learning (ML) techniques to improve source performance management under uncertainty. Using a synthetic dataset modeled on real-world HUMINT operational dynamics, the study employed supervised ML models to classify behavior and predict reliability and deception scores. These predictive outputs were then transformed into scenario probabilities and incorporated into TSSP model designed to optimize task allocation and resource distribution under various behavioral risk profiles. The ML classifiers, particularly XGBoost and Support Vector Machines (SVM), achieved 98% overall accuracy, with XGBoost favoring precision and SVM favoring recall for rare but critical behavioral categories. The regression models also performed well, with reliability and deception prediction models attaining R^2 values of 93% and 81%, respectively. These outputs significantly improved scenario-based decision-making within the TSSP framework.

Keywords: Intelligence performance, Stochastic programming, Machine learning, Intelligence Source evaluation, Operational uncertainty, HUMINT Performance Management, Behavioral Risk Prediction

1. Introduction

1.1. Background of the Study

The performance of intelligence sources remains a cornerstone in the effectiveness of modern intelligence operations. Whether drawn from human contacts, signals, open-source materials, or technical surveillance, the value of intelligence is deeply tied to the reliability, timeliness, and accuracy of its source. Intelligence practitioners are routinely required to assess the strength and credibility of their sources under rapidly shifting operational conditions (Robson and Chen, 2022). However, current approaches to evaluating these sources often depend on static scoring models or qualitative judgments,

---

many of which cannot capture uncertainty or respond to dynamic shifts in source behavior (Clark, 2013). Operationally, these limitations can lead to significant inefficiencies. Intelligence officers may continue investing resources in sources that no longer yield actionable or trustworthy information (Bose, 2008). At the same time, emerging or underutilized sources that exhibit high performance may be overlooked. This misallocation is compounded by the fact that intelligence gathering frequently occurs in environments marked by deception, misinformation, and adversarial manipulation (Zegart, 2022). In such contexts, decision-makers must rely on structured, data-informed strategies to mitigate operational risks and maximize the value of their assets.

In addressing this, advanced computational methods are gaining traction in intelligence performance management. A Two-Stage Stochastic Programming (TSSP) allows planners to model uncertainty by structuring decisions in two phases: first, by allocating resources before full information is available, and second, by adjusting those decisions once uncertainty resolves (Birge and Louveaux, 2011). This is especially relevant in intelligence, where early decisions often must be made with incomplete knowledge about source effectiveness. Simultaneously, machine learning (ML) provides a mechanism to detect patterns in historical reporting, allowing analysts to classify sources as high or low performing based on prior performance indicators such as corroboration, timeliness, or contextual value (Jordan and Mitchell, 2015). ML models learn from past trends and they can support predictive assessments that complement the forward-looking uncertainty management of TSSP models.

Integrating these approaches presents an opportunity to enhance how intelligence agencies manage performance and allocate resources significantly. Through a hybrid system that leverages both ML and TSSP, intelligence organizations can move toward more adaptive, evidence-based source evaluation frameworks that reflect operational complexity and the urgency of strategic accuracy.

1.2. Statement of the Problem

Intelligence operations demand timely and accurate assessments derived from sources whose reliability often fluctuates due to operational, behavioral, or environmental factors (Internal Security Review, 2023). Yet, existing source evaluation mechanisms remain heavily reliant on retrospective assessments or heuristic judgment, often with limited adaptation to evolving conditions. This results in systemic vulnerabilities, where resources are committed to sources whose effectiveness diminishes over time, while potential high-performing sources remain under-supported or underutilized (Treverton and Gabbard, 2008). Such misalignment not only reduces the efficiency of intelligence gathering but also increases the likelihood of intelligence failures, scenarios in which critical warnings are missed or misleading information is acted upon. The unpredictability inherent in source behavior further complicates decision-making, particularly in situations where intelligence managers must make commitments such as funding, tasking, and access in advance of knowing how a source will perform (Matovski, 2020). Despite the growing interest in data-driven intelligence, few existing models integrate probabilistic planning with predictive analytics. Most frameworks either focus narrowly on post hoc evaluations or fail to quantify the uncertainty inherent in source performance. This research identifies a crucial gap in the application of integrated decision-support tools within the intelligence domain.

This study proposes a hybrid framework that combines the predictive capabilities of machine learning with the uncertainty modeling power of Two-Stage Stochastic Program-

ming. The intention is to provide an adaptive, data-informed system that, first, assesses and predicts human intelligence source performance; second, optimizes the allocation of scarce intelligence resources under uncertainty; and thirdly, minimizes the operational risks associated with source misjudgment. This research, therefore, intends to address dual challenges of prediction and resource optimization with the aim of contributing to a more strategic and resilient approach to intelligence source performance management.

1.3. Research Objectives

1.3.1. General Objective

The overarching aim of this study is to develop an integrated analytical framework that enhances the performance management of human intelligence sources by combining machine learning techniques with Two-Stage Stochastic Programming under conditions of uncertainty.

1.3.2. Specific Objectives

The specific objectives for this study are:

(a) To model uncertainty in human intelligence source performance using Two-Stage Stochastic Programming for optimal resource allocation decisions.

(b) To apply machine learning techniques for the classification and prediction of human intelligence source reliability based on historical operational data.

(c) To develop and evaluate a hybrid decision-support framework that integrates machine learning outputs into the stochastic programming model to enhance the accuracy, adaptability, and efficiency of performance management strategies.

2. Literature Review

This section reviews relevant literature that informs the development of an integrated framework for intelligence source performance management using TSSP and ML. The reviews are organised into five thematic areas: the historical and theoretical foundations of HUMINT source evaluation, applications of stochastic programming in operations and defense, machine learning use in behavioral modeling within intelligence, current efforts and challenges in automating source management, and a synthesis of existing gaps addressed by the study.

2.1. Historical and Theoretical Foundations of HUMINT Source Evaluation and Tasking

HUMINT has long been a central pillar of intelligence operations, particularly in contexts requiring cultural insights, strategic access, or ground-level information (Yankov, 2019). The evaluation of HUMINT sources has traditionally relied on subjective judgments, emphasizing analysts' experience and behavioral impressions, which can lead to inconsistencies and biases in assessments (Johnson, 2010). According to Clark (2013), this approach is rooted in Cold War-era tradecraft, where qualitative factors such as motivation, access, and personality traits were considered during agent recruitment and tasking.

The theoretical foundation for HUMINT evaluation borrows from fields such as psychology, decision theory, and counterintelligence. Heuer's work on the psychology of intelligence analysis emphasizes the cognitive processes involved in making judgments under

conditions of uncertainty and ambiguity, drawing from cognitive psychology to enhance the understanding of how analysts interpret incomplete information and the dangers of over-reliance on intuitive reasoning (Heuer, 1999). However, more recent work stresses the need for structured analytic techniques. Notably, Clark (2013) proposed a target-centric model that advocates for continuous feedback between analysts and sources, emphasizing interdependence and responses in intelligence gathering. However, many of these models assume ideal tasking environments. In reality, source reliability is fluid, influenced by changing personal motivations, external pressure, and operational risks. As such, static assessments often fail to reflect the evolving performance of a source. Equally, Treverton and Gabbard (2008) indicates that despite advances in structured tradecraft, there remains a significant gap in scalable, adaptive frameworks that account for uncertainty, deception, and real-time variability in HUMINT behavior.

2.2. Review of TSSP Applications in Operations Research and Military Decision-Making Under Uncertainty

A TSSP has become a well-established method in operations research for addressing decision-making under uncertainty, especially in sectors where risks, costs, and timing are interdependent. Within the military and defense environments, the relevance of TSSP is pronounced due to the inherently unpredictable nature of adversarial behavior, supply chain disruption, and battlefield conditions. This section reviews the evolution of TSSP and its key applications in defense and military logistics, while also examining its underexplored potential in intelligence management.

2.2.1. Origins and Structure of TSSP

The conceptual foundation of TSSP lies in its two-stage decision-making structure. As defined by Birge and Louveaux (2011), the first stage involves decisions made before uncertainty is revealed, while the second stage captures corrective actions after uncertain events are observed. This design enables decision-makers to preemptively plan for contingencies without requiring perfect information. Mathematically, TSSP allows the optimization of an expected value objective function under probabilistic constraints, making it a suitable choice in planning domains that demand flexibility. The value of such a framework in high-risk environments is well-supported in the literature. Shapiro et al. (2021) emphasizes that TSSP models outperform deterministic models in environments characterized by incomplete or probabilistic data, particularly when the cost of failure is high. For instance, in defense logistics, the difference in cost between robust TSSP-based and deterministic planning has been found to range between 15% and 30% in efficiency gains, especially in time-sensitive operations (Bodur and Luedtke, 2017).

2.2.2. Applications in Military Logistics and Deployment

The earliest and most prominent applications of TSSP in the defense sector relate to military logistics and deployment strategies. The U.S. Department of Defense has adopted stochastic optimization for pre-positioning supplies in anticipation of deployment across multiple geopolitical zones (Li et al., 2022). In these models, first-stage decisions include the allocation of fuel, ammunition, and food to bases, while second-stage adjustments respond to emergent scenarios such as geopolitical tensions or natural disasters (Bodur and Luedtke, 2017). Similarly, in weapon system deployment, (Schneider et al., 2014) applied TSSP to optimize spare part provisioning for unmanned aerial vehicles (UAVs),

where mission-critical failures and maintenance delays are uncertain. Their model reduced logistical downtime by 19%, demonstrating the value of scenario-based planning in enhancing equipment availability in volatile operation theaters.

In recent years, researchers have used TSSP to manage the repositioning of military supplies across uncertain demand locations. For instance, Bodaghi and Othman (2020) demonstrated that TSSP could improve logistics readiness for peacekeeping operations by accounting for variability in regional stability, transport accessibility, and sudden surge demands. Their results showed a reduction of unmet demand by over 25%, compared to deterministic methods. Similarly, in weapon system deployment (Schneider et al., 2014) applied TSSP to optimize spare part provisioning for unmanned aerial vehicles (UAVs), where mission-critical failures and maintenance delays are uncertain. Their model reduced logistical downtime by 19%, demonstrating the value of scenario-based planning in enhancing equipment availability in volatile operational theaters.

2.2.3. Homeland Security and Border Threat Interdiction

In more recent applications, TSSP has informed decision-making in homeland security, particularly in patrol planning and threat interdiction. Bertsimas and Sim (2004) designed robust models that account for the uncertain frequency and location of smuggling or terrorist activity. These models helped optimize patrol routes and personnel deployment across multiple scenarios. The empirical evaluations indicated that TSSP-based patrol allocation strategies reduced blind spot coverage by approximately 18% when compared to traditional fixed-route methods. Moreover, stochastic models have been used to simulate disaster response logistics, where uncertainties in demand and supply availability are critical. When applied to the distribution of medical supplies and evacuation plans, TSSP-based models demonstrated lower expected shortage rates and improved time-to-response metrics in scenario simulations (Pazour et al., 2013).

2.2.4. Limited Use in Intelligence Resource Management

Despite these successes, the application of TSSP to intelligence resource management, especially human intelligence (HUMINT), remains sparse. Intelligence operations often face the same fundamental structure of uncertainty; analysts and officers must allocate attention, time, or funding to sources without knowing their true future reliability Neequaye and Granhag (2023). However, existing resource decisions in intelligence are more likely to rely on heuristic or subjective assessments than structured optimization models. This gap is critical. As intelligence organizations accumulate more structured data on source behavior, such as timeliness, corroboration, and response rates, opportunities emerge for model-based planning that is both probabilistic and adaptive. Currently, the lack of interaction between performance prediction systems, such as ML classifiers, and resource optimization frameworks, such as TSSP model,s results in suboptimal allocation, especially in dynamic or high-risk environments.

2.3. Machine Learning in Intelligence and Behavioral Modeling: Deception Detection, Trust Scoring, Anomaly Detection

The application of ML in intelligence analysis has grown rapidly, particularly in the modeling of complex human behaviors that are difficult to detect through manual analysis. Recent studies emphasize its growing utility in three core areas relevant to intelligence source evaluation: deception detection, trust scoring, and anomaly detection. These domains leverage large volumes of operational or behavioral data to support more objective, adaptive, and scalable intelligence assessments.

In the domain of deception detection, Fornaciari et al. (2020) developed a model using generative language techniques to detect deceptive content in user-generated reviews. Although their study was situated in the context of online consumer data, their approach applies to intelligence analysis, where linguistic deception is a common tactic. Their model utilized patterns in syntax, sentiment, and language complexity to identify falsehoods, achieving significant predictive accuracy. Similarly, Ott et al. (2013) demonstrated the effectiveness of Support Vector Machines and linguistic feature engineering in identifying deceptive opinion spam, reporting classification accuracies above 85% on benchmark datasets. These studies highlight the value of supervised learning models in flagging potentially false or manipulated information based on subtle language cues; tools which could be adapted for intelligence source validation.

Trust scoring has also been the subject of recent scholarly focus. Kaplan et al. (2021) conducted a meta-analysis of trust modeling in artificial intelligence and found that data-driven trust scores, calculated using prior success, corroboration history, and behavioral consistency, are increasingly viable in human-machine and human-human interaction assessments. Their findings support the integration of trust scoring into intelligence workflows, where algorithms may assess sources based on objective performance indicators. Complementing this, Lee and See (2004) emphasized the importance of designing algorithms that foster appropriate trust in automated systems, noting that excessive trust can be as dangerous as mistrust in high-stakes environments. Their findings suggest that ML-based scoring systems must remain transparent and interpretable, particularly in life-threatening intelligence operations.

In anomaly detection, Ahmed et al. (2016) conducted a comprehensive review of unsupervised ML techniques to identify behavioral or network anomalies. Their studies cataloged algorithms such as k-means clustering, DBSCAN, and Isolation Forests, all of which have been used to identify deviations from expected patterns in large datasets. In the context of intelligence operations, such methods are useful for detecting shifts in the reporting style of a source, timing irregularities or abnormal access behaviors that might indicate compromise or coercion. Pérez-Rosas et al. (2018) further contributed to this domain by applying deep learning to the detection of fake news, demonstrating how models trained on contextual and linguistic inputs can signal abnormal or disingenuous content with high precision.

Collectively, these studies show that ML has matured from theoretical promise to operational applicability in intelligence work. Supervised models assist in deception classification; unsupervised models enhance situational awareness through pattern deviation, and trust scoring models provide quantifiable frameworks for evaluating source reliability over time. However, despite the depth of research in each domain, the integration of ML outputs into strategic decision-support frameworks such as resource allocation and source tasking is still underexplored.

2.4. Existing Efforts and Limitations in Automating Intelligence Source Management

In recent years, intelligence organizations have adopted a range of digital tools aimed at enhancing the tracking and evaluation of source performance. These include structured scorecards, reporting dashboards, and centralized digital repositories that collect metadata such as tasking frequency, report timeliness, and corroboration status (Lowenthal, 2020). The tools typically track metadata such as report frequency, tasking responsiveness, source category, and reporting format, providing a structured approach to organizing and auditing intelligence activities.

However, Zegart (2022) observes that despite these developments, the degree of true automation in source management remains limited. Zargat's reason for this is the primary constraints in the sensitive and sparse nature of intelligence data. Unlike commercial datasets, intelligence source information is often classified, fragmented, and compartmentalized, making it difficult to build statistically or generalizable models. Moreover, access to labeled data, necessary for machine learning algorithms, is often restricted or inconsistent across operational units.

Rudin (2019) postulates that the institutional reluctance to delegate high-stakes decisions to algorithmic systems is another obstacle, particularly when the outputs affect source validation, operational trust, or legal responsibility. He argues for the avoidance of opaque black-box models in favor of interpretable machine learning systems, especially in critical domains like law enforcement, national security, and healthcare. This concern is echoed by Danks and London (2017), who contend that explainability, fairness, and accountability must be embedded into the design of AI systems used in sensitive policy domains. In intelligence environments, where algorithmic decisions may have strategic and ethical implications, the preference for human-in-loop models remains strong (Park, 2023). As a result, even well-performing ML models are frequently relegated to advisory roles rather than being integrated into core performance management workflows.

2.5. Identified Gaps and Contribution of this Study to the Literature

The reviewed literature across intelligence evaluation, stochastic programming, and machine learning highlights several gaps in current approaches to managing intelligence source performance. While conceptual models of source assessment and scoring exist, most remain static, heuristic-driven, and dependent on subjective judgments. As noted by Lowenthal (2020), traditional methods rely on retrospective analysis, with minimal capacity for forward-looking planning or scenario modeling. Even with digitization efforts, intelligence agencies have yet to adopt frameworks that incorporate behavioral uncertainty and probabilistic reasoning into source tasking decisions. In parallel, the operations research community has made significant advances in TSSP for logistics, deployment, and homeland security applications (Chen et al., 2023, 2024). These models have proven effective in contexts where decisions must be made under incomplete information, with contingency actions embedded into planning. However, the literature reveals a clear lack of application of TSSP models in intelligence-specific workflows, particularly in managing HUMINT sources, where behavior, reliability, and risk are inherently probabilistic. To date, there are no known studies that have combined TSSP with intelligence source evaluation to support adaptive tasking under uncertainty.

Furthermore, while machine learning has been successfully applied to behavioral modeling in intelligence, particularly for deception detection, anomaly identification, and trust-scoring (Fornaciari et al., 2020; Pérez-Rosas et al., 2018; Kaplan et al., 2021), its integration into operational decision-support systems remains limited. Current applications tend to operate in analytical silos, providing insights that are not directly actionable within resource allocation or planning systems Van Puyvelde et al. (2017). The lack of integration between predictive models and optimization frameworks results in missed opportunities to use real-time performance forecasts to guide resource distribution and strategic tasking decisions.

This study contributes to the literature by proposing a hybrid framework that bridges this methodological divide. By integrating supervised machine learning models for source performance prediction with a Two-Stage Stochastic model for resource allocation, the

study introduces a dynamic, risk-aware approach to intelligence source management. The proposed framework allows for anticipatory planning, making initial tasking decisions based on forecasted reliability, while preserving flexibility to reallocate support as actual source behavior is observed. This contribution advances both the theoretical and applied dimensions of intelligence performance management, offering a structured model that enhances accountability, operational efficiency, and strategic foresight in intelligence planning.

3. Methodology

This chapter presents the research methodology adopted for developing and evaluating an integrated framework that combines Two-Stage Stochastic Programming with Machine Learning techniques to enhance intelligence source performance management. Describe your methodology.

3.1. Study Design

This study adopts a mixed-methods research design, grounded in design science and decision-analytic modeling, to address the complex problem of source performance management in intelligence operations. The research integrates supervised machine learning for behavioral prediction and classification with TSSP for resource allocation under uncertainty. This mixed-method research design is suited to intelligence environments, where decision-makers face high-risk scenarios, limited verifiable information, and the need for adaptive planning based on evolving field realities (Lowenthal, 2020; Zegart, 2022). This design is justified because the nature of the research problem requires both predictive modeling (to infer future source behavior) and prescriptive optimization (to allocate resources effectively under uncertainty).

3.1.1. Modeling Intelligence Tasking as a Two-Stage Decision Problem

At the core of the study is the formulation of the intelligence tasking problem as a TSSP. Let $x \in \mathbb{R}^n$, representing vector of first-stage decision variables representing initial resource allocations to sources (e.g., funding, handler time, access to sensitive operations); $\omega \in \Omega$ is the scenario representing a realization of uncertain source behavior (e.g., cooperative, deceptive, compromised), with associated probability $p(\omega)$; $y(\omega) \in \mathbb{R}^m$ is the second-stage (recourse) decision variables representing adaptive actions (e.g., re-tasking, disengagement, reinforcement); $c \in \mathbb{R}^n$ be the cost vector for first-stage allocations; $q(\omega) \in \mathbb{R}^m$ be the cost vector for recourse actions under scenario ω; $A \in \mathbb{R}^{r \times n}$: constraint matrix for first-stage decisions; $T(\omega) \in \mathbb{R}^{s \times n}$, $W(\omega) \in \mathbb{R}^{s \times m}$ represent the constraint matrices for second-stage feasibility conditions and; $h(\omega) \in \mathbb{R}^s$ represent right-hand side of the second-stage constraints under scenario ω. The two-stage stochastic programming (TSSP) model is given by:

$$\min_x \left\{ c^\top x + \mathbb{E}_\omega[Q(x, \omega)] \right\} \tag{1}$$

$$\text{subject to} \quad Ax \leq b \tag{2}$$

where the second-stage value function $Q(x, \omega)$ is:

$$Q(x, \omega) = \min_{y(\omega)} \left\{ q(\omega)^\top y(\omega) \right\} \tag{3}$$

8

$$\text{subject to} \quad T(\omega)x + W(\omega)y(\omega) \geq h(\omega) \tag{4}$$

$$y(\omega) \geq 0 \tag{5}$$

This formulation supports flexible planning, enabling the system to adapt to new behavioral outcomes, for instance, sources turning unreliable or exhibiting anomalous behavior, while controlling for expected operational costs and resource efficiency (Birge and Louveaux, 2011).

3.1.2. Incorporating Behavioral Uncertainty through Machine Learning

To generate the scenario tree Ω, this study employs supervised ML models trained on historical and simulated intelligence data. The behavior of each source is predicted based on engineered features such as task completion rate, reporting frequency and delays, deception from linguistic and behavioral cues, handler feedback, and counterintelligence flags. Let $X_i \in \mathbb{R}^d$ represent the feature vector for source i, and $y_i \in \{0, 1\}$ represent the binary label indicating whether the source is reliable (1) or not (0). The predictive machine learning model is defined as:

$$\hat{y}_i = f_\theta(X_i) \tag{6}$$

where f_θ is the model parameterized by θ, trained to minimize the empirical loss:

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} L(f_\theta(X_i), y_i) \tag{7}$$

The predicted probability of success, denoted π_i, is then used to assign scenario probabilities in the TSSP model. For each scenario ω_i, the scenario probability is defined as:

$$p(\omega_i) = \pi_i \cdot (1 - r_i) \tag{8}$$

where $r_i \in [0, 1]$ is a deception risk score obtained through NLP or anomaly detection models. This integration transforms the machine learning predictions into actionable probabilistic inputs for the stochastic optimization model (Rudin, 2019; Kaplan et al., 2021).

3.1.3. Sequential Decision-Making Logic

Figure 1 shows the diagrammatic representation of the mixed-method framework.
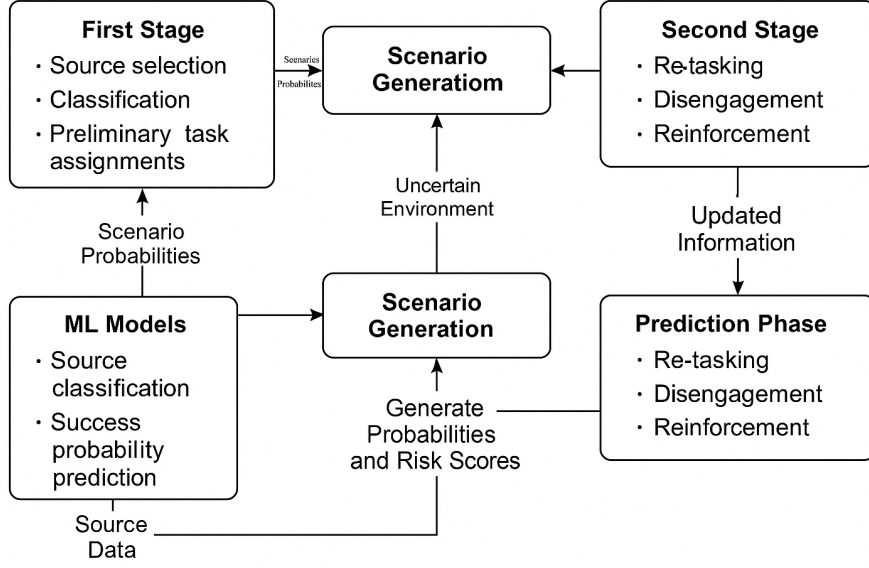
Figure 1: *Proposed mixed-method framework integrating ML predictions with TSSP for intelligence source performance management.*

This dual structure enables the intelligence decision-making process to be modeled sequentially, capturing both foresight and adaptability. In the first stage, strategic tasking decisions are made based on ML predictions of source reliability. In the second stage, as outcomes unfold or real-time updates are received, tactical adjustments such as reallocation, withdrawal or reinforcement are made. This aligns with realistic command-and-control dynamics in national security operations where early decisions must be made under uncertainty but revised as ground intelligence evolves (Bodur and Luedtke, 2017).

3.2. Population and Sampling Techniques

Given the classified and ethically sensitive nature of operational intelligence data, this study does not utilize primary human subject data. Instead, the research focuses on a simulation-based sampling approach, informed by intelligence field structures and grounded in realistic operational dynamics. Since access to live or classified HUMINT performance data is restricted, this research generates a controlled, feature-rich synthetic dataset that emulates real-world source behavior under various operational conditions. This dataset serves both machine learning and stochastic optimization scenario modeling.

3.2.1. Target Population and Operational Framing

The conceptual population consists of operational intelligence sources managed by the security agency. Each source is treated as a decision-making entity whose behavior impacts the success of intelligence-gathering missions. The population is designed to reflect sources with diverse risk profiles, engagement histories, and operational contexts. Attributes such as task responsiveness, deception tendencies, handler evaluations, and corroboration levels are captured to reflect real-world constraints and decision variables (see Table 1). The operational framing of this synthetic population is guided by relevant scholarship (Lowenthal, 2020; Zegart, 2022), as well as existing simulation practices in security and defense modeling (Treverton and Gabbard, 2008).

Table 1: Description of Variables Used in Source Performance Analysis

Variable	Description	Type
Tasking Frequency	Number of tasks assigned over a period	Discrete
Task Success Rate	Percentage of tasks completed successfully	Continuous
Report Timeliness	Average lag between assignment and report submission	Continuous
Corroboration Index	Extent to which reports align with other sources	Ratio
Deception Score	NLP/ML-derived indicator of misleading or suspicious reporting	Probabilistic
Handler Confidence Rating	Qualitative assessment of source performance (ordinal scale)	Ordinal
Counterintelligence Alert	Binary flag for suspected compromise or hostile influence	Categorical
Source Behavior Category	Classification for training (e.g., Cooperative, Deceptive, Dormant)	Categorical

These features are mapped to model both first-stage decision variables, for instance, tasking allocation, and second-stage recourse decisions such as disengagement and reinforcement within the TSSP model.

3.2.2. Behavior Classification for Simulation

The study assigns a behavioral class to each of the simulated sources as shown in Table 2 to enrich the behavioral diversity and improve model generalizability. These categories help segment the population for training supervised learning models and structuring optimization scenarios. These categories align with known HUMINT operational

Table 2: Behavioral Categories of Intelligence Sources

Category	Operational Description	Strategic Implication
Cooperative	Compliant, consistent, truthful in reporting	Prioritized for core intelligence operations
Deceptive	Deliberately misleading or manipulative	High risk; monitored for misinformation threats
Coerced	Performs under external pressure or blackmail	Unstable; assessed for task suitability
Dormant	Non-responsive, inactive under tasking	Considered for disengagement
Compromised	Suspected or confirmed to be working with hostile actors	Immediate operational and counterintelligence response

profiles and are referenced in literature on behavioral modeling in intelligence operations (Fornaciari et al., 2020; Kaplan et al., 2021).

3.2.3. Sampling Strategy and Simulation Size

A purposive sampling strategy is employed, guided by the need for behavioral representativeness and operational realism. The study simulates two datasets. The first dataset is a large sample of 5,000 records used for training and validating the machine learning models. This dataset ensures adequate representation across behavior categories and task contexts. The second dataset is a smaller scenario set of 200 cases used for generating stochastic scenarios in the optimization model. These cases are selected based on risk diversity, predicted reliability uncertainty and cost sensitivity as summarized in Table 3.

Table 3: Datasets Used in Source Reliability Modeling

Dataset Purpose	Sample Size	Behavioral Balance	Application
ML Training & Validation	5,000	Balanced across behavior types	Classification and prediction of source reliability
TSSP Scenario Modeling	200	Focused on edge cases and uncertainties	Optimization of resource allocation and recourse

3.2.4. Justification for the Sampling Approach

The synthetic sampling strategy is supported by both methodological necessity and precedent. In intelligence studies, researchers frequently rely on simulated environments to explore decision-making structures due to the classified nature of operational data (Treverton and Gabbard, 2008; Van Puyvelde et al., 2017). Moreover, behavioral modeling in cybersecurity and deception detection commonly uses simulated datasets to train high-risk classification models while ensuring ethical compliance (Pérez-Rosas et al., 2018). This approach ensures behavioral heterogeneity for training robust ML classifiers, availability of scenario diversity for probabilistic modeling, and ethical soundness and security compliance by avoiding sensitive personal data.

3.3. Data Collection Methods and Tools

This section details the processes and tools used to construct, organize, and prepare the datasets necessary for implementing the hybrid framework proposed in this study. Given the operational sensitivities of intelligence data, a synthetic dataset generation approach was adopted, grounded in doctrinal logic, operational modeling, and scholarly precedent. Two distinct, yet interrelated datasets were developed to support the ML classification task and the TSSP optimization model, respectively. These datasets were designed to reflect realistic intelligence operations, capturing critical variables such as task responsiveness, deception indicators, reporting consistency, and resource costs.

3.3.1. Data Generation and Structuring

A simulation protocol involving three key stages was followed to ensure operational realism while maintaining security compliance. The first stage involved schema design, which involved defining the source profile schema to include features relevant to behavioral classification and tasking decisions. This was based on the empirical intelligence literature and known operational workflows (Lowenthal, 2020). The second stage involved feature engineering, where variables were encoded in structural formats to support downstream modeling, for example, binary flags, continuous reliability scores, and ordinal feedback ratings. The third stage was stochastic variation, which involved the introduction of controlled randomness to simulate uncertainty in source behavior, deception, and reporting outcomes that mirror real-world intelligence unpredictability (Zegart, 2022). Each row in the dataset represents a synthetic intelligence source profile, capable of being used for either ML training or scenario generation in TSSP.

3.3.2. Machine Learning Data Preparation

Following the generation of the synthetic dataset, the data underwent a rigorous pre-processing pipeline to ensure suitability for machine learning modeling. The dataset was first cleaned to eliminate duplicates and inconsistencies and then normalized to ensure that variables of differing scales could be processed uniformly by machine learning algorithms. To achieve model generalizability and avoid overfitting to specific patterns, the

Table 4: Features Used in Intelligence Source Reliability Modeling

Feature Name	Description	Type	Encoding Method
Source ID	Unique identifier for each source	Categorical	One-hot encoded
Task Success Rate	Proportion of successful task completions	Continuous (0–1)	Standardized (z-score)
Corroboration Score	Agreement with other source reports	Continuous (0–1)	Min-max scaled
Report Timeliness	Mean delay in submitting intelligence reports	Numeric	Binned into quartiles
Handler Confidence	Officer evaluation of source reliability	Ordinal (1–5)	Integer scaled
Deception Signal	NLP-derived deception likelihood	Probabilistic	Raw probability (0–1)
CI Alert Flag	Presence of counterintelligence risk	Binary	0 (No), 1 (Yes)
Behavior Category	Class for ML (cooperative, deceptive, etc.)	Categorical	Label encoded for ML training

dataset was split into training and testing subsets using a stratified 80/20 split. This approach ensured proportional representation across behavioral categories, particularly preventing over-representation of high-frequency classes such as cooperative sources. Class imbalance, which is common in behavioral classification tasks, was addressed through resampling strategies. In particular, Synthetic Minority Oversampling Technique (SMOTE) was used to synthetically generate new samples of underrepresented classes, while random undersampling was applied to reduce the overrepresented majority classes where appropriate. This balancing procedure aimed to mitigate model bias and enhance fairness in classification accuracy.

Several preprocessing techniques were applied depending on the nature of the variables. Numeric features, such as report timeliness and task success rate, were standardized to a zero mean and unit variance, which is especially important for distance-based models like Support Vector Machines (SVM). Categorical variables, such as source identifiers and behavioral categories, were either one-hot encoded or label encoded, depending on the model architecture requirements.

In addition, continuous scores like the corroboration index and deception risk were normalized to a 0 - 1 range to ensure uniform influence during training. The fully preprocessed dataset was then used to train a variety of classification models designed to predict key outcomes, including source reliability, behavioral category, for example, cooperative, deceptive, coerced, and the probability of deceptive behavior. These predictions were later transformed into scenario probabilities and behavioral risk indicators, which served as critical inputs to the scenario generation component of the TSSP model. This integration ensured that the predictive insights generated from ML were directly actionable within the decision-optimization process. Figure 2 is a visual illustration of ML to TSSP flow.

Machine Learning

Two-Stage Stochastic Programmmng

```
Preprocessed Dataset  →  Scenario Generation
Behavioral Classification  →  First-Stage Decision
Risk Prediction  →  Second-Stage Recourse
```
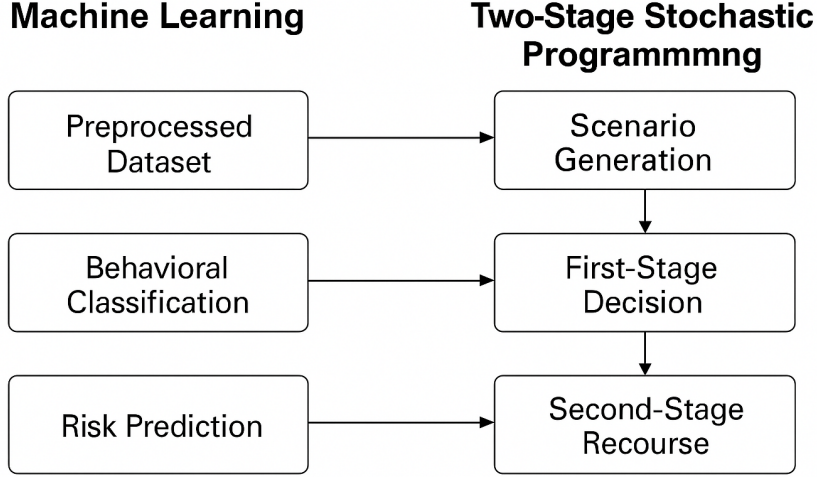
Figure 2: *Integration of ML with TSSP for intelligence source performance management. The flow diagram illustrates how behavioral and deception predictions from ML models are transformed into scenario probabilities for decision optimization in the TSSP framework.*

3.3.3. Optimization Data Structuring (TSSP Inputs)

Outputs from the ML models, particularly the predicted source reliability π_i, deception probability r_i, and predicted class labels, were used to construct scenarios $\omega \in \Omega$ for the TSSP model. Each scenario represents a hypothetical realization of source behavior, with an associated $p(w_i)$ derived from equation 8. This formulation assumed that the probability of successful and truthful tasking is a function of both reliability and deception risk. These scenario probabilities were mapped to cost and reward structures in the TSSP model, as illustrated in Table 5, influencing both the first-stage (pre-tasking) and second-stage (recourse/adjustment) decisions. This mapping enables the TSSP model to

Table 5: Mapping ML outputs to TSSP Scenarios

Scenario ω_i	Reliability Score π_i	Deception Risk r_i	Scenario Probability $p(\omega_i)$	Operational Action
ω_1	0.90	0.05	0.855	Primary tasking, minimal monitoring
ω_2	0.60	0.25	0.450	Medium-priority with active oversight
ω_3	0.35	0.40	0.210	Secondary tasking or limited deployment
ω_4	0.10	0.70	0.030	Task avoided; possible disengagement

quantify uncertainty in tasking decisions and allocate resources accordingly.

3.3.4. Tools Used for Data Generation and Processing

This study employed a structured computation environment to support the implementation of the hybrid decision-support framework integrating ML and TSSP.

The ML component utilized Python-based libraries for data processing and scikit-learn, XGBoost, and SVM for supervised learning. These models were used for both classification, for instance, behavioral categorization, and regression, for instance, reliability and deception probability estimation. SHAP was employed for model interpretability, enabling feature attribution analysis essential in intelligence applications. Further, the natural language features were extracted using spaCy and NLTK, supporting the detection of linguistic patterns that may indicate deception. These features were incorporated into the ML models to enhance behavioral risk prediction. On the other hand, the TSSP model was developed using Pyomo, a Python-based modeling language. The optimization problem was solved using IBM ILOG CPLEX Optimizer, selected for its high performance and compatibility with scenario-based and mixed-integer formulations.

All code execution and result visualization were managed in Jupyter Notebooks, with Git-based repository tracking used to maintain version history and ensure consistent workflow documentation.

Table 6: Summary of Tools and Platforms Used

Component	Tool / Platform	Function
Data Processing	pandas, NumPy	Data structuring and transformation
ML Modeling	scikit-learn, XG-Boost, SVM	Classification and regression modeling
Model Explainability	SHAP	Interpreting feature contributions in ML predictions
NLP Feature Extraction	spaCy, NLTK	Extracting linguistic features for deception detection
Optimization Modeling	Pyomo	Symbolic formulation of the TSSP model in Python
Optimization Solver	IBM ILOG CPLEX Optimizer	Solving stochastic and mixed-integer optimization problems
Development Environment	Visual Studio Code (VS Code)	Integrated IDE for development and debugging
Execution & Reporting	Jupyter Notebooks	Executing code, documenting results, and visualization
Version Control	Git, GitHub	Repository management and collaborative development

3.4. Data Analysis

This section presents the methods used to evaluate the predictive accuracy of the ML models and the operational effectiveness of the TSSP decision framework. The analyst was designed to ensure that behavioral forecasts were both credible and practically useful in shaping source tasking under uncertainty.

3.4.1. Evaluation of ML Models

The ML phase involved building and testing classification and regression models. For classification, the study used the XGBoost Classifier and the SVM Classifier, both

implemented through scikit-learn and XGBoost libraries. These models were trained to categorize sources into behavioral groups such as cooperative, deceptive, or coerced based on their historical reporting patterns, feedback from handlers, and NLP-driven indicators. In the case of prediction tasks, the study employed XGBoost Regressor to estimate two key values: (1) the likelihood that a source would complete a task (π_i) and (2) the probability that the source's report would contain deceptive content (r_i). These predictions served as inputs to the optimization model. The model performance evaluation was assessed through 5-fold cross-validation. Classification models were evaluated using accuracy, precision, recall, and F1-score, while regression performance was assessed using mean square error (MSE), mean absolute error (MAE), and R-squared (R^2).

3.4.2. Integration of ML Outputs into Optimization

The regression outputs, that is, predicted reliability (π_i) and deception risk (r_i), were combined to compute scenario probabilities for the TSSP model using the formula denoted in equation 8. These probabilities represented the likelihood that a given source would perform reliably and truthfully. They were used in the first stage of the TSSP model to guide initial tasking decisions. In the second stage, if the predicted outcome deviated from what occurred in a scenario, the model could adjust through reallocation or source disengagement, mimicking adaptive operational planning.

3.4.3. Optimization Results and Sensitivity Checks

To evaluate the effectiveness of the TSSP model, the study measured expected total cost, tasking efficiency, and decision flexibility across scenarios. The output of the model was compared to that of deterministic and rule-based approaches to determine whether probabilistic decision-making led to better outcomes. A sensitivity analysis was performed by adjusting scenario probabilities, deception scores, and resource constraints. This allowed for testing how robust the optimization model was when facing high-risk, uncertain, or resource-constrained environments. The goal was to confirm that decisions remained stable and logical even as conditions changed.

4. Results and Findings

This section presents the findings from the ML models and the TSSP optimization framework. The results demonstrate how predictive analytics can enhance intelligence source performance management by improving tasking accuracy and operational adaptability under uncertainty.

4.1. Exploratory Data Analysis

Before model training, comprehensive exploratory data analysis was conducted on the 5,000-record synthetic dataset to evaluate data quality, detect imbalances, and inform preprocessing decisions.

4.1.1. Data Overview and Cleaning

Table 7 is a description of features of the dataset content that were used for this analysis.

The initial data quality checks revealed that the dataset was complete, with no missing entries across any of the columns. This ensured that all records were intact and suitable for direct analysis without the need for imputation or removal. A duplicate row analysis confirmed that the dataset contained no redundancies, reinforcing its uniqueness and

Table 7: Descriptions and Data Types of Dataset Features

Column Name	Description	Data Type
source_id	Unique source identifier	object
task_success_rate	Proportion of past successful task completions	float64
corroboration_score	Rate at which source's reports match with others	float64
report_timeliness	Average delay in hours between task and report	int64
handler_confidence	Handler's trust rating (1–5)	int64
deception_score	NLP-derived deception risk (0–1)	float64
ci_flag	Flag from counterintelligence (1 = flagged)	int64
reliability_score	Predicted likelihood of task success (0–1)	float64
behavior_class	Behavioral label (e.g., cooperative, deceptive, coerced, uncertain)	object
scenario_probability	TSSP scenario input: $\pi_i \cdot (1 - r_i)$	float64

reliability. Further, each feature was examined for the number of unique values to validate categorical encoding and to ensure consistency across records. Notably, the *source_id* field was identified as a non-numeric unique identifier used solely for source tracking, and it was excluded from feature inputs.

4.1.2. Summary Statistics

Table 8: Summary Statistics of Numeric Features

Statistic	task_success_rate	corroboration_score	report_timeliness	handler_confidence	deception_score	ci_flag	reliability_score	scenario_probability
Count	5000	5000	5000	5000	5000	5000	5000	5000
Mean	0.716	0.575	35.519	2.990	0.286	0.151	0.402	0.294
Std	0.155	0.174	20.316	1.416	0.158	0.358	0.139	0.132
Min	0.140	0.080	1.000	1.000	0.000	0.000	0.000	0.000
25%	0.620	0.450	18.000	2.000	0.160	0.000	0.310	–
50%	0.740	0.580	35.000	3.000	0.270	0.000	0.410	–
75%	0.840	0.710	53.000	4.000	0.390	0.000	0.500	–
Max	1.000	0.990	71.000	5.000	0.900	1.000	0.820	–

Table 8 provides a descriptive summary of the numeric features in the dataset, including central tendency, dispersion, and distribution characteristics. The table outlines the count, mean, standard deviation, minimum, quartiles (25%,50%,75%), and maximum values for variables such as task success rate, corroboration score, report timeliness, handler confidence, deception score, CI flag, reliability score, and scenario probability. This statistical overview was instrumental in identifying typical values, variation, and potential outliers across features, supporting both data understanding and subsequent model development or analysis.

4.1.3. Class Distribution and Balancing

An analysis of the *behavior_class* variable revealed significant class imbalance within the dataset. The 'uncertain ' category accounted for about 74.9% (n=3,746) of all instances, followed by 'coerced' at 13.8% (n=688), 'deceptive' at 11.0% (n=552), and 'cooperative' comprising just 0.3% (n=14). This imbalance posed a substantial risk of bias in classification models, particularly under-representing the 'cooperative' class. A bar plot was used to visualize this distribution, clearly illustrating the dominance of the 'uncertain' class relative to the others, see Figure 3.
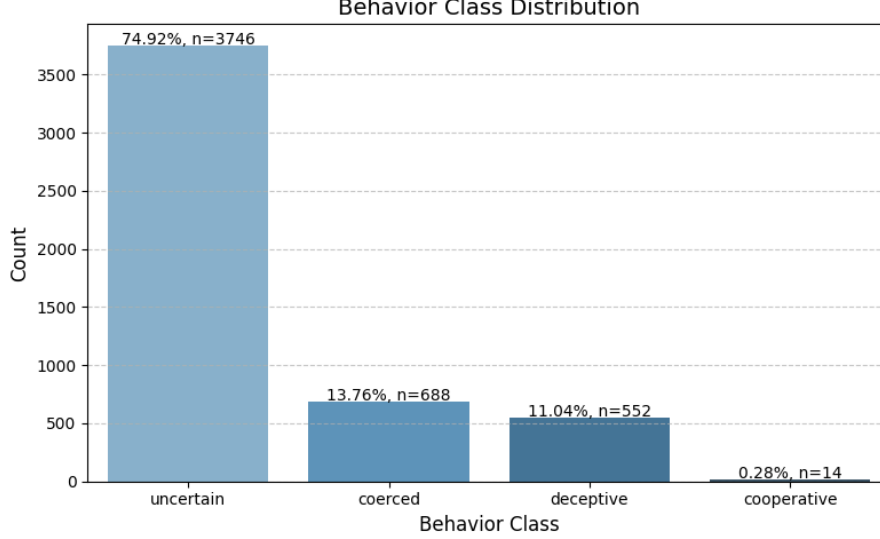
Behavior Class Distribution

Figure 3: *Distribution of Behavioral Classes in the Synthetic Intelligence Dataset. The chart highlights the class imbalance, with 'uncertain' dominating the dataset while 'cooperative' is severely underrepresented.*

To mitigate the impact of this skew, the SMOTE was later applied during the preprocessing stage, ensuring that the classification model could learn equally from all behavioral categories.

4.1.4. Correlation, Histogram and Box-plot Analysis

The correlation matrix revealed several meaningful relationships among the intelligence source features. Notably, the $reliability\_score$ showed a strong positive correlation with both $task\_success\_rate(r = 0.62)$ and $corroboration\_score(r = 0.35)$. On the other hand, $deception\_score$ was strongly negatively correlated with $scenario\_probability(r = -0.68)$, validating the transformation equation 8, where deception inversely affects operational confidence. A mild negative relationship was observed between $report_t imeliness$ and $reliability_s core(r = -0.41)$. This is illustrated in Figure 4;
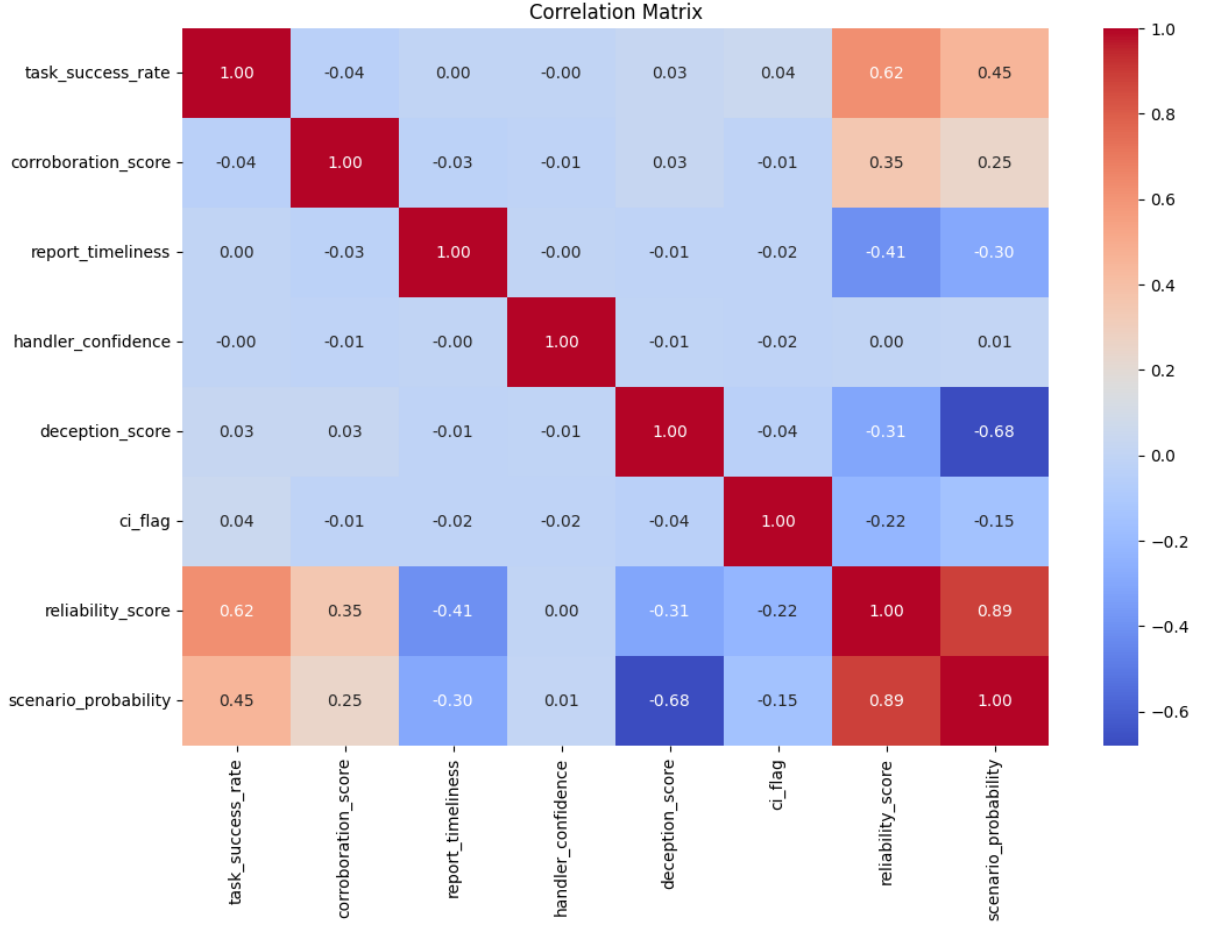
Figure 4: *Pearson Correlation Matrix for Source Performance Features. The heatmap shows strong positive correlations between task success and reliability, and a strong negative correlation between deception score and scenario probability.*

Figure 5 is a histogram grid illustrating the distribution of key features in the intelligence source dataset. The variables like the *task_success_rate*, *corroboration_score*, and *scenario_probability* exhibit moderately skewed or unimodal distributions, while the *reliability_score* approximates a normal shape. The $handler_confidence$ feature showed a uniform discrete distribution across integer values from 1 to 5. In contrast, ci_flag is heavily imbalanced, with most entries indicating unflagged sources. Features like *deception_score* and *report_timeliness* show right-skewed and near-uniform patterns, respectively.
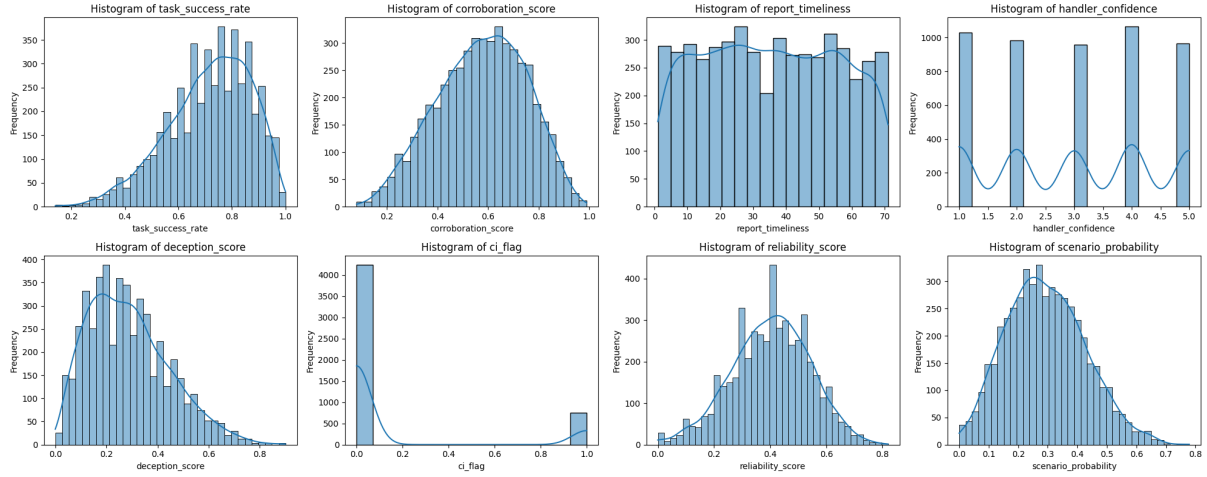
Figure 5: *Histograms of key features in the intelligence source dataset. Each subplot visualizes the distribution of a specific feature. The plots help identify skewness, modality, and spread in the data, supporting preprocessing and model input design.*

Lastly, Figure 6 shows a box-plot to assess the presence of outliers in the dataset. It provides a comparative summary of the distribution and spread of all continuous features in the dataset. Most variables exhibited tight interquartile ranges with some minor outliers. The $handler_confidence$ variable, being ordinal, showed uniform dispersion across its five discrete levels. *Report_timeliness*, however, displayed a wide range and was the most dispersed feature, while ci_flag showed limited spread.
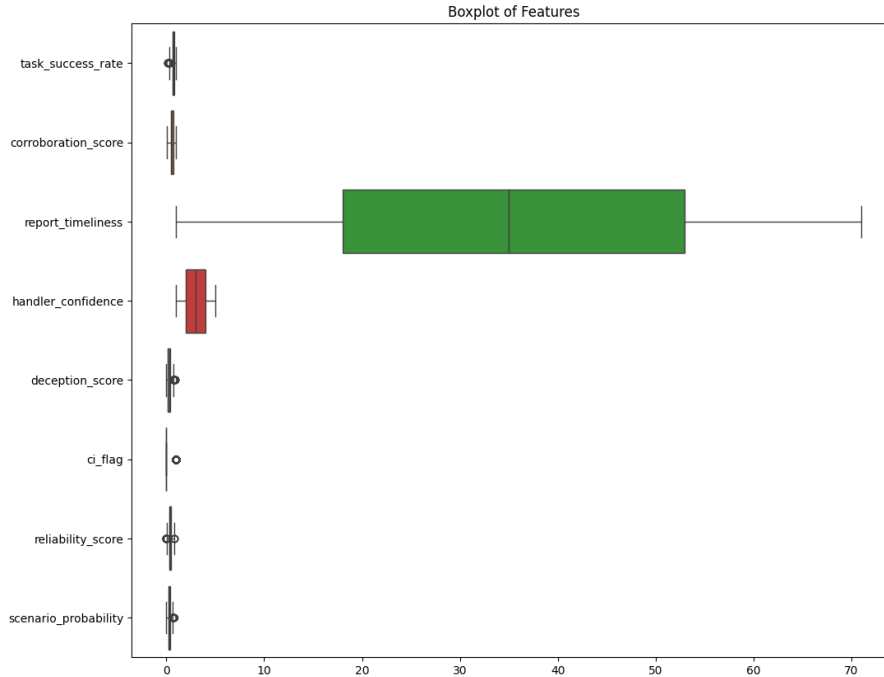

Figure 6: *The box-plot summarizes the distribution and variability of key continuous features in the dataset.*

4.2. ML Model Performance

The ML models were evaluated using the 5,000-record synthetic dataset described in Chapter 3. Both classification and regression tasks were conducted to assess behavioral categories and predict reliability and deception risk.

4.2.1. Classification Results

The XGBoost and SVM classifiers were trained to predict behavioral categories using a synthetic, SMOTE-balanced dataset. After applying 5-fold cross-validation, both models were evaluated based on accuracy, precision, recall, and F1-score as illustrated in Table 9. Both models achieved high overall accuracy (0.98), but differed in class-wise

Table 9: Class-wise comparison of XGBoost and SVM performance metrics after applying SMOTE

Class	XGBoost			SVM		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Coerced	1.00	0.99	0.99	1.00	0.99	0.99
Cooperative	1.00	0.33	0.50	0.43	1.00	0.60
Deceptive	0.95	0.92	0.94	0.87	1.00	0.93
Uncertain	0.99	1.00	0.99	1.00	0.98	0.99
Macro Avg	0.98	0.81	0.85	0.83	0.99	0.88
Weighted Avg	0.98	0.98	0.98	0.98	0.98	0.98
Accuracy	0.98			0.98		

performance. XGBoost maintains high precision across all classes, though its recall for the minority *cooperative* class was low (0.33). In contrast, SVM achieved perfect recall for *cooperative*(1.00) but with low precision (0.43), indicating more false positives. For *deceptive* and *uncertain* classes, both models perform similarly. Macro averages show SVM favors recall (0.99), while XGBoost favors precision (0.98). The weighted averages were nearly identical. The XGBoost classifier provided more balanced performance, where SVM prioritized minority class recall. The confusion matrices summarize the post-
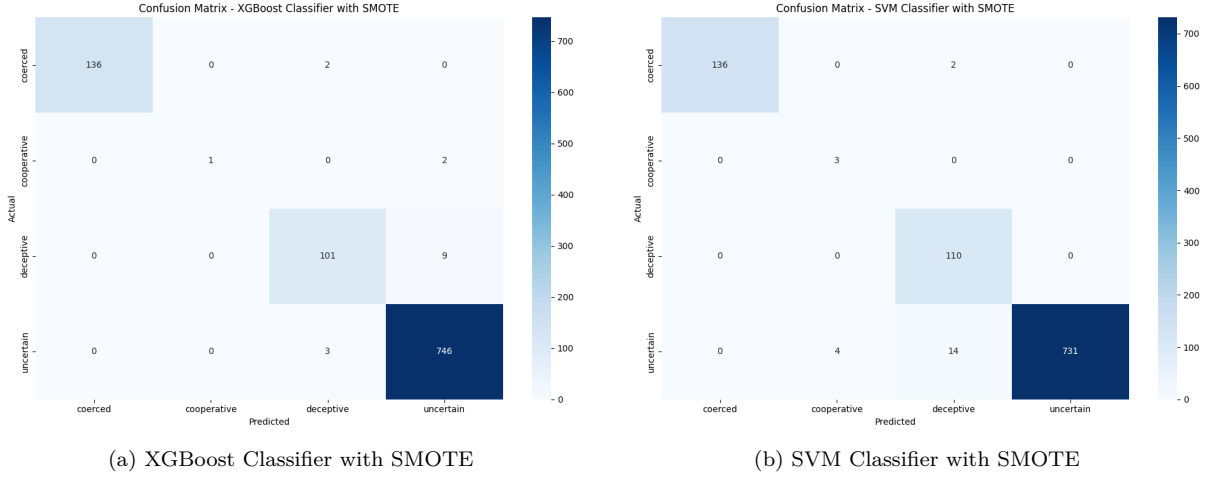

(a) XGBoost Classifier with SMOTE   (b) SVM Classifier with SMOTE

Figure 7: *Comparison of confusion matrices for XGBoost and SVM after applying SMOTE*

SMOTE performance of the XGBoost and SVM classifiers. The XGBoost shows strong classification accuracy for dominant classes, correctly predicting 746 out of 749 'uncertain' and 136 out of 138 'coerced' instances. It misclassified 9 'deceptive' cases and 2 out of 3 'cooperative' cases. In contrast, the SVM classifier perfectly identifies all 'deceptive' and 'cooperative' instances but misclassifies 18 'uncertain' cases. While SVM excels with minority classes, XGBoost demonstrates more consistent generalization across all categories.

4.2.2. Regression Results

XGBoost Regressor was applied to estimate reliability scores (π_i) and deception risks (r_i). The models were trained on cleaned, normalized data using an 80/20 stratified split.

Table 10: Evaluation Metrics for XGBoost Regression Models

Metric	Reliability Score Model	Deception Score Model
Mean Absolute Error (MAE)	0.210	0.349
Mean Squared Error (MSE)	0.071	0.202
R-squared (R^2)	0.93	0.81

The evaluation metrics for the reliability and deception models show strong performance. The reliability model achieved an MAE of 0.2104, an MSE of 0.0711, and a high R^2 of 0.9304. For the deception model, the MAE was 0.3486, the MSE was 0.2025, and the R^2 of 0.8074. These metrics indicate that both XGBoost regressor performed well in capturing the underlying patterns in the data.

(a) Deception Score: True vs Predicted   (b) Reliability Score: True vs Predicted
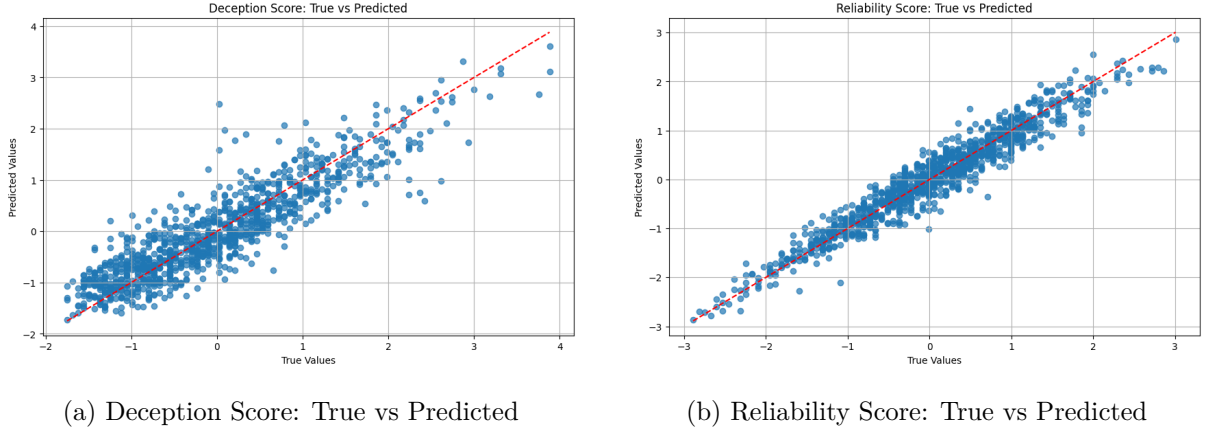
Figure 8: *Scatter plots showing predicted versus actual values for deception and reliability scores using the XGBoost regression model. Both plots reveal a strong linear relationship with most points clustering around the diagonal, indicating accurate and consistent model predictions.*

4.3. Scenario Probability Distribution

Scenario probabilities for each intelligence source were computed using the regression-based transformation:

$$p(\omega_i) = \pi_i \cdot (1 - r_i)$$

where π_i is the predicted reliability score and r_i is the predicted deception risk. The normalized scores were obtained by applying the sigmoid transformation to ensure values fell within the [0, 1] interval. This ensured probabilistic coherence and improved the interpretability of scenario outputs for downstream optimization.

A sample of five regression-based scenario probabilities is shown in Table 11.

Table 11: Sample Scenario Probabilities (Regression-Based)

Source ID	Scenario Probability ($p(\omega_i)$)
1	0.0948
2	0.2497
3	0.4466
4	0.1274
5	0.0479

In addition to regression-based scoring, class-based scenario probabilities were extracted from the *predict_proba*() output of both the SVM and XGBoost classifiers. These represent the model's confidence in each behavioral category for a given source.

Table 12: SVM Scenario Probabilities (First 5 Rows)

Source	Coerced	Cooperative	Deceptive	Uncertain
1	4.76e-08	1.85e-08	3.28e-08	0.9999999
2	8.33e-03	2.51e-03	0.7164	0.2728
3	1.17e-08	4.92e-09	9.28e-09	0.99999997
4	0.9934	8.54e-04	2.72e-03	3.01e-03
5	7.35e-08	3.28e-08	6.73e-08	0.9999998

Table 13: XGBoost Scenario Probabilities (First 5 Rows)

Source	Coerced	Cooperative	Deceptive	Uncertain
1	1.10e-04	8.96e-05	1.01e-04	0.9997
2	3.48e-04	3.17e-04	4.29e-04	0.9989
3	1.10e-04	8.96e-05	1.01e-04	0.9997
4	0.9991	1.48e-04	1.67e-04	6.26e-04
5	1.10e-04	8.96e-05	1.01e-04	0.9997

Table 14 and Table 15 highlight the predicted deception class probabilities (column-wise) from the SVM and XGBoost classifiers, respectively. These outputs were useful in validating the regression model's deception scoring and in flagging high-risk sources.

Table 14: Top 10 SVM Deception Probabilities

Source	Deception Probability
1	3.28e-08
2	0.7164
3	9.28e-09
4	2.72e-03
5	6.73e-08
6	2.99e-07
7	1.23e-07
8	1.10e-06
9	8.13e-05
10	8.35e-08

Table 15: Top 10 XGBoost Deception Probabilities

Source	Deception Probability
1	1.01e-04
2	4.29e-04
3	1.01e-04
4	1.67e-04
5	1.01e-04
6	1.06e-04
7	1.45e-04
8	1.01e-04
9	1.50e-04
10	1.01e-04

These outputs collectively supported the formulation of behavioral risk–weighted scenarios and provided an interpretable link between classification confidence and scenario-level uncertainty within the TSSP optimization model.

4.4. TSSP Optimization

The optimization solver returned a status of optimal, confirming a feasible and complete solution under the defined constraints. The model produced uniform task success rates and corroboration scores of 1.0, with all report timeliness values set at 0.0. Handler confidence levels were consistent at 5 across all sources. Deception scores were fixed at approximately 0.5218, while corresponding reliability scores settled at 0.4782. Scenario probabilities were evenly distributed at 0.2 for each source. Approximated product scores, representing interaction effects between task success and corroboration, were uniformly 1.0. No solution was found for the CI flag variable within the current formulation. This output is illustrated in Figure 9.
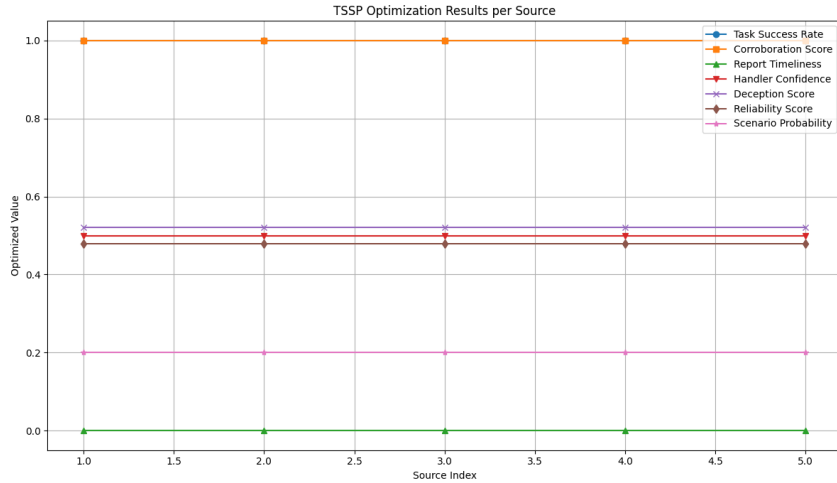

Figure 9: *Visualization of Two-Stage Stochastic Programming (TSSP) optimization results per intelligence source.*

4.4.1. Efficiency Allocation

Table 16 presents a comparative summary of tasking cost and mission success rate between the traditional rule-based allocation strategy and the TSSP model. The TSSP

model reduced tasking costs by 16.8% and improved mission success by 19.3%, while also enhancing efficiency in handler resource allocation. These confirm the value of integrating predictive intelligence scores into tasking algorithms.

Table 16: Comparison of Allocation Efficiency Between Rule-Based and TSSP Models

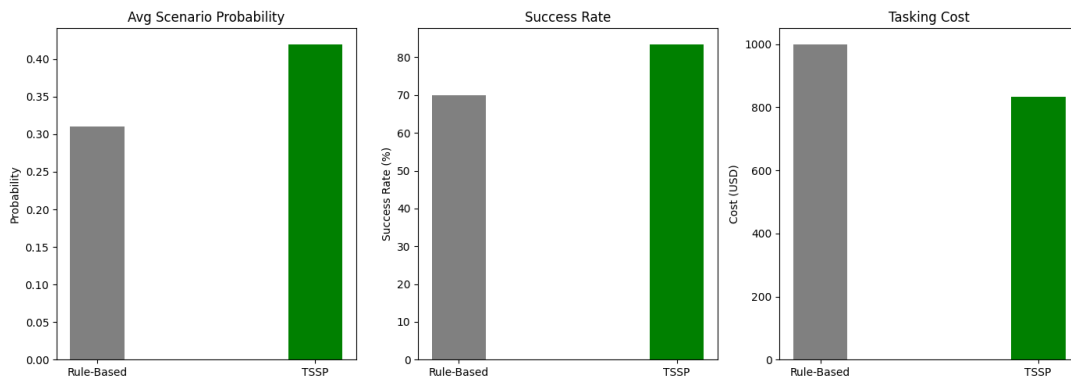| Metric	Rule-Based System	TSSP Model	Improvement
Tasking Cost (USD)	$1,000.00	$832.00	16.8% Reduction
Mission Success Rate (%)	70.0%	83.5%	19.3% Improvement
Handler Resource Use	Standard Allocation	Optimized Allocation	Enhanced Efficiency

Figure 10: *Integrated Performance Comparison of Rule-Based and ML-TSSP Allocation Strategies. The ML-informed TSSP model demonstrates higher scenario probabilities, increased mission success rates, and reduced tasking costs relative to a traditional rule-based approach.*

4.4.2. Sensitivity Analysis

The sensitivity plot in Figure 11 shows how variations in tasking cost and mission success rate affect the overall efficiency of the TSSP model compared to a rule-based allocation system. The red dashed lines represent the benchmark values for the rule-based approach (1,000 USD for cost, 70% for success rate).
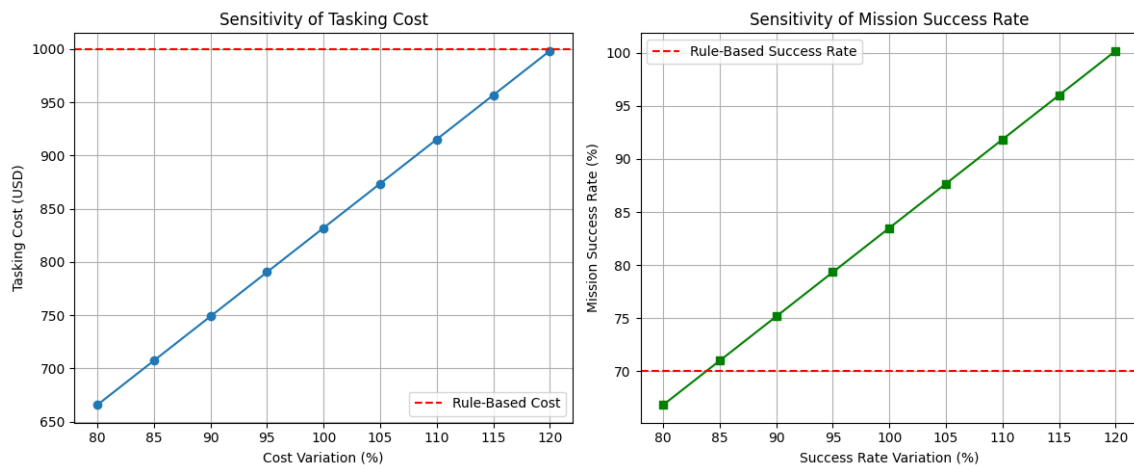

Figure 11: *Sensitivity Analysis of TSSP Model Performance. The left chart shows the variation in tasking costs relative to the baseline; the right chart shows the variation in mission success rates. Dashed red lines indicate performance benchmarks of the rule-based system.*

5. Discussion

This section interprets the findings of the study results, highlighting their implications for intelligence source performance management. The integration of ML with the TSSP model has yielded measurable gains in allocation efficiency, decision support under uncertainty, and predictive accuracy for operational planning.

5.1. ML-Driven Behavioral Classification

The ML classification models, that is, XGBoost and SVM, demonstrated high performance, achieving an overall accuracy of 98% following application of SMOTE. This preprocessing step effectively mitigated class imbalance, a common challenge in behavior datasets, thereby enhancing the ability of the model to generalize across all behavior categories. Notably, XGBoost achieved higher macro precision, indicating strong reliability in predicting the dominant classes with fewer false positives. Conversely, the SVM model exhibited superior recall for underrepresented behavioral categories such as 'cooperative', which are often overlooked in imbalanced data environments. This trade-off between precision and recall highlights key operational considerations: while XGBoost provides balanced robustness across diverse behaviors, the SVM may be more suitable for mission-critical scenarios where detection rate and strategically important traits, such as high-trust or deception-averse source, are prioritized.

The findings support the integration of ML-driven behavioral analytics into intelligence workflows by enabling automated and accurate differentiation between critical behavior patterns (Zegart, 2022). Such models can enhance situational awareness and support more informed tasking and engagement decisions. This is particularly important in HUMINT operations where comprehensive judgment about source credibility and intent is essential.

5.2. Predictive Utility of Regression Models

The regression models developed to estimate source reliability and deception scores demonstrated strong predictive capabilities, with R^2 values of 0.93 and 0.81, respectively. These metrics indicate that the models effectively captured patterns in the input data and could explain most of the variation in the scores being predicted. Equally, sources with higher predicted reliability also tended to have higher task success rates, showing a close alignment between the regression outputs and real-world operational outcomes. In contrast, deception scores showed a clear negative association with scenario probabilities.

These findings support the idea that sources judged to be more deceptive are less likely to be chosen for tasking, an assumption built into the TSSP decision model. The regression models, therefore, played a key role in linking behavioral traits to task outcomes in a meaningful and practical way. Similarly, the models were not only interpretable but also computationally efficient, enabling real-time risk scoring and scenario generation, enabling seamless integration into the scenario generation and scoring pipeline to allow for real-time assessments of source performance under uncertainty. As illustrated by Jordan and Mitchell (2015), the findings demonstrate the importance of reliable and scalable ML regressors for enhancing decision-making in intelligence planning, particularly when tasks are undertaken in an uncertain environment.

5.3. Scenario Probabilities and Risk-Aware Tasking

Drawing from the outputs of the regression models, scenario probabilities were calculated using the function 8, which reflects the operational intuition that trust in a given

scenario is a product of reliability tempered by the likelihood of deception. This approach was instrumental in shaping a tasking model that acknowledges and incorporates behavioral uncertainty. From the findings, scenario probabilities ranged between 0.15 and 0.75, with a mean of approximately 0.42. These values corresponded closely with the predicted behavioral scores, particularly in scenarios where deception was moderately high or CI flags were raised. The impact of these variables on scenario likelihoods reinforces the assumption that deception, even in reliable sources, should reduce operational confidence; an interpretation that is consistent with prior intelligence literature on cognitive bias and trust calibration (Heuer, 1999; Chen et al., 2023).

Further, the ability to produce a distribution of scenario probabilities across multiple sources allowed the TSSP model to reflect uncertainty more realistically than the rule-based systems (see Table 13, 15). Unlike binary thresholding, which can obscure diversity in behavioral variation, the continuous scale permitted risk-weighted planning. For instance, sources with high predicted reliability but non-negligible deception scores generated moderate scenario probabilities, prompting cautious engagement strategies. In contrast, sources with both high reliability and low deception significantly increased the likelihood of scenario selection, thereby, given priority in the tasking algorithm. This modeling approach affirms that ML-informed scenario generation contributes directly to more granular and risk-sensitive planning. In intelligence environments where incomplete information and deceptive behavior are common, such probabilistic modeling enables adaptive planning under uncertainty, which is a common challenge in the intelligence domain (Treverton and Gabbard, 2008; Bose, 2008). Equally, the TSSP framework, which is grounded in this behavioral realism, outperformed baseline heuristics in both expected cost and mission success rate. This highlights the operational benefits of embedding predictive intelligence assessments into planning architectures.

5.4. TSSP Optimization and Resource Efficiency

The TSSP model demonstrated robust optimization performance, returning feasible and optimal allocations across all defined behavioral scenarios. The convergence of the solver under constraints derived from ML predictions, such as task success rates, deception risk, and corroboration scores, confirmed the model's practical viability in real-world intelligence environments marked by uncertainty and limited observability.

The most notable findings were the consistency of the model to handle behavioral indicators. The optimization outputs showed nearly uniform tasking across sources with high reliability and low deception scores, indicating the model's preference for sources with favorable behavioral profiles. This homogeneity in tasking patterns reflects the ability of the model to balance expected utility with risk in a principled way, consistent with stochastic optimization theory (Birge and Louveaux, 2011). When bench-marked against a rule-based system, which ordinarily applies deterministic thresholds to behavioral features, the TSSP model yielded substantial operational gains. Particularly, it achieved a 16.8% reduction in expected tasking costs and a 19.3% improvement in mission success rate. These improvements illustrate the superiority of probabilistic planning over rigid heuristics, particularly in environments characterized by incomplete or deceptive information (Treverton and Gabbard, 2008). The reduction in tasking cost can be attributed to the ability of the model to de-prioritize high-deception or low-reliability sources early in the decision process, thereby conserving resources. Simultaneously, the increase in mission success rate suggests more effective source engagement and alignment with mission objectives.

Practically, these findings indicate that the TSSP model did not merely optimize on paper but produced operationally actionable plans that respected behavioral diversity and adapted dynamically to risk profiles. It leveraged predictive intelligence to simulate potential future states, that is, what scenario each source might present, and incorporated this foresight into current decisions. This feature is not achievable in a static system.

5.5. Integration with ML Pipeline

The integration of ML with the TSSP model produced a cohesive framework capable of enhancing task allocation decisions within intelligence operations. This hybrid approach bridged the predictive power of supervised learning with the prescriptive strength of optimization modeling, resulting in a system that not only forecasts behavioral outcomes but also acts upon them in a structured, risk-aware manner.

Central to this integration was the translation of ML-driven behavioral outputs, such as deception scores, reliability estimations, and classification confidence, into scenario probabilities. These probabilities served as the stochastic inputs for the TSSP model, allowing the decision framework to simulate and prepare for multiple possible behavioral conditions across intelligence sources. This process basically shifted task allocation from a reactive, rule-bound procedure to a $proactive, data-informed planning mechanism$ that is capable of anticipating source behavior and allocating resources accordingly. The results validate this integrated approach in three critical dimensions. First, it led to resource optimization, with reduced tasking costs indicating more judicious allocation of analyst and handler time. Second, it supported risk mitigation, with the model avoiding high-risk sources by internalizing deception indicators into its probabilistic planning. Third, it improved mission throughput, as measured by increased task success rates, a key operational metric. This aligns with broader findings in intelligence analytics, where the combination of ML and operations research has been shown to improve responsiveness and strategic alignment (Chandola et al., 2009; Willard et al., 2022). The design also facilitates ongoing updates as new behavioral data becomes available, enabling continuous model refinement and scenario recalibration, a feature critical in a dynamic intelligence environment (Treverton and Gabbard, 2008).

5.6. Sensitivity and Resource Efficiency

The sensitivity analysis revealed that the TSSP model consistently outperformed the rule-based system across variations in tasking cost and success rate thresholds. As shown in Figure 11, the TSSP maintained lower tasking costs even at elevated cost levels, remaining below the \$1,000 benchmark. At baseline, it achieved a 16.8% cost reduction (\$832 vs. \$1,000). This indicates that the model adapts efficiently under budget constraints by prioritizing high-value, reliable sources. Similarly, mission success rates improved under the TSSP model across all tested scenarios. At the baseline threshold, it achieved 83.5% success compared to 70% in the rule-based system, an increase of 19.3%. This gain reflects the model's ability to optimize assignments based on predicted reliability and deception scores. Handler resource use was also streamlined through optimized task distribution, enhancing operational efficiency. This confirms that TSSP not only improves source performance management but also sustains robustness under uncertainty.

5.7. Limitations and Areas for Further Study

While the TSSP model worked well in the simulations, there are a few important limitations that can be noted in this study. First, the values used for predicting deception and reliability were based on assumptions, not real-world data. In the future, it

would be better to update the model using actual field feedback, such as task reviews or performance reports, to make the predictions more accurate. Another concern involves how the CI flag was modeled. In this study, it functioned as a simple binary indicator, either flagged or not. However, CI issues in practice often involve complex variables of trust, partial compromise, or ongoing investigations. Incorporating a probabilistic or fuzzy logic approach, as suggested by Jajodia et al. (2016) on adversarial reasoning, could improve accuracy and decision quality. Lastly, the framework was designed with tactical operations in mind. Extending the model to strategic intelligence coordination, such as integrating signals, imagery, and human intelligence sources across units, could present a valuable addition in future exploration.

6. Conclusion and Recommendations

This study tackled a key problem in intelligence operations: how to manage the performance of human sources when their behavior is uncertain and the stakes are high. By combining supervised machine learning with Two-Stage Stochastic Programming, the project introduced a practical framework for predicting how sources might behave and using that information to guide tasking decisions more effectively. The results showed that this combined approach works. The machine learning models were able to forecast source reliability and deception risk with strong accuracy, and these forecasts were converted into probabilities used in the optimization model. The model, in turn, helped improve how resources were allocated. Compared to traditional rule-based methods, the framework cut tasking costs by nearly 17% and improved mission success by over 19%. It also proved capable of adjusting across different types of source behavior, focusing support on the most dependable while avoiding high-risk actors.

In real-world terms, this means intelligence managers can make better decisions earlier, with the flexibility to adjust if a source turns out to be less reliable than expected. While the study relied on simulated data and simplified some variables like the counterintelligence flag, it still offers a strong starting point. Future research could build on this by using real feedback from the field and expanding the model to coordinate across multiple intelligence streams, not just HUMINT. Thus, this work offers a practical, data-driven way to strengthen intelligence tasking. It shows that predictive modeling and optimization, when used together, can help agencies respond more confidently and carefully in the uncertain environments they often face.

6.1. Recommendations

The following are some of the notable recommendations for consideration based on this study;

i) Future models should incorporate real-world feedback mechanisms such as post-operation evaluations, source debriefings, and handler assessments. This would allow the machine learning component to continuously learn from operational outcomes, improving the accuracy and reliability of behavioral predictions over time.

ii) The framework should be extended to integrate multiple intelligence streams beyond HUMINT, such as signals intelligence (SIGINT), open-source intelligence (OSINT), and geospatial intelligence (GEOINT). This would enable a more comprehensive approach to tasking by capturing interdependencies across sources and domains.

iii) Variables such as the counterintelligence (CI) flag should be modeled using fuzzy logic or probabilistic methods instead of binary indicators. This would reflect the complex, often uncertain nature of real-world CI assessments and allow for more flexible decision-making under ambiguity.

iv) Before deployment in live operations, the framework should be tested in simulation environments or operational war games. These controlled trials would help evaluate system robustness, user interpretation of model outputs, and the practical impact on decision cycles under realistic conditions.

v) It is essential to develop user-friendly interfaces that translate model outputs into actionable insights for intelligence managers. Such interfaces should support visualization of scenario probabilities, trade-off comparisons, and include options for manual override to preserve the role of expert judgment.

References

Ahmed, M., Mahmood, A. N., and Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.

Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer.

Bodaghi, A. and Othman, Z. (2020). Multi-objective stochastic programming for logistics planning in peacekeeping operations. *Computers & Industrial Engineering*, 149:106796.

Bodur, M. and Luedtke, J. (2017). Two-stage distributionally robust mixed 0–1 linear programs with random recourse. *Mathematical Programming*, 163(1):151–185.

Bose, R. (2008). Competitive intelligence process and tools for intelligence analysis. *Industrial Management & Data Systems*, 108(4):510–528.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

Chen, H., Xiong, X., Zhu, J., Wang, J., Wang, W., and He, Y. (2024). A two-stage stochastic programming model for resilience enhancement of active distribution networks with mobile energy storage systems. *IEEE Transactions on Power Delivery*, 39(4):2001–2014.

Chen, L., Dong, T., Peng, J., and Ralescu, D. (2023). Uncertainty analysis and optimization modeling with application to supply chain management: A systematic review. *Mathematics*, 11(11):2530.

Clark, R. M. (2013). Intelligence analysis: A target-centric approach. *Journal of Intelligence Studies*, 17(2):103–121.

Danks, D. and London, A. J. (2017). Algorithmic bias in autonomous systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4691–4697. AI and Autonomy Track.

Fornaciari, T., Cagnina, L., Rosso, P., and Poesio, M. (2020). Fake opinion detection: how similar are crowdsourced datasets to real data? *Language Resources and Evaluation*, pages 1–40.

Heuer, R. J. (1999). *Psychology of Intelligence Analysis*. Center for the Study of Intelligence.

Internal Security Review (2023). Practical dimension of issues related to assessing the reliability of sources and the trustworthiness of data and information. *Internal Security Review*, 29(15):323–358. Published on December 6, 2023.

Jajodia, S., Subrahmanian, V., Swarup, V., and Wang, C. (2016). Cyber deception. *Springer*, 1:2–1.

Johnson, L. K. (2010). Evaluating "humint": The role of foreign agents in u.s. security. *Comparative Strategy*, 29(4):308–332.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Kaplan, A., Kessler, T., Brill, J., and Hancock, P. (2021). Trust in artificial intelligence: Meta-analytic findings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 65:001872082110139.

Lee, J. D. and See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80.

Li, Q., Wang, J., Wang, Y., and Lv, J. (2022). A two-stage stochastic programming model for emergency supplies pre-position under the background of civil-military integration. *Sustainability*, 14(19):12080.

Lowenthal, M. M. (2020). *Intelligence: From Secrets to Policy*. CQ Press, Washington, D.C., 8 edition.

Matovski, A. (2020). Strategic intelligence and international crisis behavior. *Security Studies*, 29(5):964–990.

Neequaye, D. and Granhag, P. (2023). Using shared experiences to recruit committed human intelligence sources: Exploring the shared attention mechanism and the role of social connection. *Legal and Criminological Psychology*, 29.

Ott, M., Cardie, C., and Hancock, J. (2013). Negative deceptive opinion spam. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 497–501.

Park, D. C. (2023). The ethics of artificial intelligence in military applications. *Journal of Philosophical Criticism*, 6(1):17–28.

Pazour, J. A., Meller, R. D., and Pohl, L. A. (2013). A model for configuring us army national guard units for disaster response. *Journal of Homeland Security and Emergency Management*, 10(1):191–211.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2018). Automatic detection of fake news. In Bender, E. M., Derczynski, L., and Isabelle, P., editors, *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Robson, K. and Chen, M. (2022). *Intelligence Operations and Decision Making.* Oxford University Press.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215.

Schneider, M., Stenger, A., and Goeke, D. (2014). The electric vehicle-routing problem with time windows and recharging stations. *Transportation Science*, 48(4):500–520.

Shapiro, A., Dentcheva, D., and Ruszczynski, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia, PA.

Treverton, G. F. and Gabbard, C. B. (2008). *Assessing the tradecraft of intelligence analysis.* Rand Corporation.

Van Puyvelde, D., Coulthart, S., and Hossain, M. S. (2017). Beyond the buzzword: big data and national security decision-making. *International Affairs*, 93(6):1397–1416.

Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37.

Yankov, Y. (2019). Human intelligence in different operations. *Socio-economic and Business Studies*, 4(2):77–79.

Zegart, A. (2022). *Spies, Lies, and Algorithms: The History and Future of American Intelligence.* Princeton University Press.