JSEG

JOURNAL OF SMART ECONOMIC GROWTH

www.jseg.ro ISSN: 2537-141X Volume 11, Number 2, Year 2026

ANALYSING THE DETERMINANTS OF GRADUATE UNEMPLOYMENT IN TUNISIA USING MACHINE LEARNING

Sami Mestiri

Faculty of Management and Economic Sciences of Mahdia, University of Monastir, Tunisia.

Rue Ibn Sina, Hiboun, Mahdia, Tunisia.

https://orcid.org/0000-0002-2060-3242

Abstract: *This article analyses the determinants of youth graduate unemployment in Tunisia by combining classical econometric methods (logistic regression) with three machine learning algorithms (Random Forest, XGBoost, RBF-kernel SVM) applied to an original survey of 1,200 Tunisian graduates. The econometric results reveal that female gender, belonging to the engineering field, and education employment mismatch are the most significant determinants. The machine learning analysis confirms the predominance of gender in discriminating between unemployed and employed individuals, and uncovers non-linear relationships that parametric models fail to capture. XGBoost and SVM offer the best predictive performance. These findings call for a deep reform of the university system, targeted policies against gender discrimination, and improved recruitment transparency.*

JEL classification: J64, J24, C25, C45

Keywords: graduate youth unemployment, Tunisia, logistic regression, random forests, XGBoost, education–employment relationship, human capital.

1. INTRODUCTION

For two decades, Tunisia has exhibited a remarkable socio-economic paradox: despite a spectacular expansion of higher education the enrolment rate increased from 12% in 1990 to over 36% in 2020 graduate youth unemployment has remained chronically high, oscillating between 28% and 35% depending on the year. This phenomenon, referred to as the diploma paradox, reflects a structural disconnect between the supply of university training and the actual needs of the labour market (Mestiri, 2025a; Mestiri, 2025b).

The revolution of 14 January 2011 brought to light the demands for employment and dignity from an overqualified yet labour-market-excluded youth. According to the National Institute of Statistics (INS, 2023), the national unemployment rate stood at 15.2% in 2023; however, higher-education graduates recorded a rate of 26.7%, compared with only 4.5% for individuals without qualifications. Our survey of 1,200 graduates confirms and amplifies this diagnosis: the observed unemployment rate reaches 30.9%, with considerable gender disparities.

This study aims to identify and rank the determinants of unemployment by mobilising a dual approach: econometric (logit/probit) and machine learning (Random Forest, XGBoost, SVM). The originality of this work lies in the application of these approaches to a sample of 1,200 graduates, providing sufficient statistical power to detect effects that prior small-sample studies could not identify.

Three hypotheses structure the analysis:

- H1: Education-employment mismatch is the primary determinant of unemployment.
- H2: Female gender significantly increases the probability of unemployment.
- H3: Reliance on personal connections is the most discriminating factor according to machine learning.

The remainder of the article is organised as follows: Section 2 presents the theoretical framework; Section 3 describes the data and methodology; Section 4 presents the empirical results; Section 5 discusses the policy implications; Section 6 concludes.

2. THEORETICAL FRAMEWORK

2.1 Human Capital and Signalling Theories

Human capital theory, formalised notably by Gary Becker (1964) and Jacob Mincer (1974), rests on the idea that education constitutes a productive investment. It improves skills, raises the marginal productivity of workers, and translates, on the labour market, into higher wages and a lower probability of unemployment. Within this framework, the expected relationship is monotonically decreasing: the higher the level of education, the lower the risk of unemployment.

However, our empirical results suggest a weakening, if not a breakdown, of this mechanism in Tunisia. The data confirm that master's degree holders (28.7%) exhibit a lower unemployment rate than bachelor's degree holders (32.8%), yet the level of education is broadly not statistically significant across our models, suggesting a generalised devaluation of credentials. This absence of statistical significance indicates that educational investment no longer systematically translates into a differential advantage on the labour market — a form of devaluation of formal human capital linked to education–employment mismatch and the saturation of the skilled labour market (World Bank, 2020; ILO, 2022).

In this context, the signalling approach proposed by Michael Spence (1973) offers a complementary reading: a degree is less a factor of productivity than a signal allowing employers to infer unobservable characteristics (discipline, cognitive ability, motivation). When higher education was elitist, the degree constituted a strong and credible signal. Conversely, the massification of university access in Tunisia has led to credential inflation, reducing their power of differentiation.

This idea is extended by Kenneth Arrow (1973) through screening theory. According to this approach, the educational system acts as a sorting mechanism allowing employers to rank individuals rather than genuinely increasing their productivity. In a context of massification, this mechanism becomes less efficient: degrees no longer allow fine discrimination among candidates, pushing employers to rely on other selection criteria.

2.2 Matching Models and Frictions

The matching models arising from job search theory, notably developed by Mortensen and Pissarides (1994), analyse unemployment as the result of frictions on the labour market. In contrast to the neoclassical instantaneous equilibrium approach, these models introduce a matching function linking the number of unemployed workers and the number of vacancies, emphasising that the meeting between labour supply and demand is costly and imperfect.

These frictions take several forms: job search costs, information asymmetries between employers and candidates, skill mismatch, and institutional rigidities. They imply that even in the presence of vacancies, persistent unemployment can subsist due to matching difficulties.

In the Tunisian context, these mechanisms are particularly relevant. The concentration of employment opportunities in Greater Tunis creates a spatial fragmentation of the labour market. Graduates from interior regions face high mobility costs (financial, social, and informational), limiting their access to available jobs. This situation generates a structural imbalance in which high unemployment and unfilled vacancies coexist, reflecting inefficiency in the matching process (World Bank, 2020).

Furthermore, information asymmetries are reinforced by the low transparency of the labour market and the role of informal networks in recruitment. Firms may struggle to assess candidate quality, while young graduates lack reliable information on available opportunities. This dual informational deficit extends matching delays and increases unemployment duration.

2.3 Social Capital and Professional Networks

The analysis of the role of social networks in access to employment draws on the work of Mark Granovetter (1973), who introduced the concept of the "strength of weak ties." Counterintuitively, it is not close relationships (family, intimate friends) that are most effective in finding a job, but more distant connections, because they provide access to new and non-redundant information about professional opportunities.

This framework is situated more broadly within the notion of social capital, developed notably by Pierre Bourdieu (1986), who defines it as the set of resources accessible through networks of relationships. Social capital then acts as a complement — or even a substitute — for human capital, by facilitating access to information, reducing uncertainty for employers, and accelerating the matching process on the labour market.

Our empirical results strongly confirm the relevance of these approaches in the Tunisian context. The pervasiveness of nepotistic hiring ("pistonnage") reflects a marked reliance on personal networks in recruitment processes: 82% of respondents consider that personal connections play an important or very important role in obtaining employment (Q15, mean score 4.25/5). This finding suggests that the labour market is strongly relational and that access to employment depends largely on an individual's position within relevant social networks.

From an econometric standpoint, the importance of this variable is corroborated by the machine learning models, in which indicators related to social networks emerge as major discriminating variables between unemployed and employed individuals. This implies that, all else equal, individuals endowed with higher social capital have a significantly greater probability of accessing employment.

This predominance of networks can nonetheless be interpreted ambivalently. On the one hand, it reduces certain informational frictions by facilitating the circulation of information on job openings. On the other hand, it may generate inefficiencies and inequalities by favouring co-optation logic over criteria of competence and merit. In this sense, the widespread reliance on informal networks can be viewed as a symptom of an institutional deficit in the labour market, where formal matching channels (public employment services, recruitment platforms) do not play their full role.

2.4 The Contribution of Machine Learning

Since the seminal work of Leo Breiman (2001) and more recent developments such as those of Tianqi Chen (2016), supervised machine learning algorithms have profoundly renewed predictive analysis in the social sciences. These approaches random forests, boosting, neural networks are distinguished from traditional econometric models by their capacity to handle complex data structures.

Unlike parametric econometric models (logit, probit), which rely on explicit functional assumptions (linearity, additivity, error distribution), machine learning models are flexible and non-parametric. They automatically capture non-linear relationships, threshold effects, and interactions among variables that are often difficult to specify a priori within a classical econometric framework. This flexibility is particularly useful in unemployment analysis, where determinants are multiple, interdependent, and potentially heterogeneous across individuals.

Moreover, these methods are optimised for prediction rather than causal inference. They aim to minimise out-of-sample classification or prediction error using techniques such as cross-validation and regularisation. In this context, Chalfin et al. (2016) show that machine learning models can outperform traditional econometric approaches in certain decision-making tasks, notably in candidate evaluation and resource allocation.

Applied to our study, machine learning provides significant added value. First, it identifies the most important variables in predicting unemployment through measures such as feature importance. Second, it improves the overall predictive performance of the model by capturing complex structures present in survey data. For example, the high importance of variables related to social capital or labour market perceptions suggests that the determinants of unemployment are not exclusively individual, but also relational and contextual.

However, this predictive power comes with certain limitations. Machine learning models are often described as "black boxes" because their economic interpretation is less direct than that of econometric models. This poses a challenge in terms of understanding underlying mechanisms and formulating policy recommendations. Accordingly, a complementary approach combining econometrics and machine learning — sometimes referred to as a dual culture — appears particularly well suited to reconciling interpretability and predictive performance.

JOURNAL OF SMART ECONOMIC GROWTH

www.jseg.ro ISSN: 2537-141X Volume 11, Number 2, Year 2026

3. DATA AND METHODOLOGY

3.1 Data Source and Sample

The study draws on a questionnaire survey administered between February and March 2025 to 1,200 young Tunisian graduates (n = 1,200), constituting a convenience sample targeting diversity of profiles by gender, level of education, field of study, and employment situation.

The questionnaire comprises 23 questions distributed across five sections: sociodemographic profile (Q1–Q6), labour market perceptions (Q7–Q11, Likert scale 1–5), obstacles to employment (Q12–Q15), education–employment adequacy (Q16–Q20), and open-ended questions (Q21–Q23).

Table 1. Sample composition (n = 1,200)

Variable	Category	Count	%
Gender	Male	648	54.0
	Female	552	46.0
Age	< 25 years	360	30.0
	25–30 years	576	48.0
	> 30 years	264	22.0
Education level	Bachelor's	624	52.0
	Master's	528	44.0
	Doctorate	48	4.0
Field of study	Economics/Management	528	44.0
	Engineering	336	28.0
	Sciences	264	22.0
	Other	72	6.0
Current situation	Unemployed	371	30.9
	Employed	264	22.0
	Student	240	20.0
	Internship	216	18.0
	Self-employed	96	8.0
Internships completed	Yes	852	71.0
	No	348	29.0
Training adequacy	Yes	312	26.0

	No	504	42.0
	Partially	384	32.0

3.2 Dependent Variable and Explanatory Variables

The dependent variable Y is binary: Y = 1 if the individual is unemployed, Y = 0 otherwise. With 371 unemployed out of 1,200 respondents, the observed unemployment rate is 30.9%.

The explanatory variables include: gender (binary, female = 1), age (dummy variables: 25–30 years, over 30 years), level of education (dummy variables: master's, doctorate), field of study (dummy variables: engineering, sciences), education–employment adequacy (dummy variables: no, partially), absence of internships (binary), and labour market perception scores.

3.3 Econometric Methods

Logistic Model

The logistic model estimates the conditional probability of being unemployed as a function of a vector of explanatory variables Xi. The retained specification is:

$$P(Yi = 1 \mid Xi) = 1 / [1 + \exp(-(\beta_0 + \beta^T Xi))] \quad (1)$$

where Yi = 1 indicates unemployment and Xi encompasses all individual and socioeconomic characteristics. The parameters βk are estimated by maximum likelihood. Coefficient interpretation relies on odds ratios, defined as $\exp(\hat{\beta} k)$, which measure the multiplicative effect of a unit change in the explanatory variable on the odds of being unemployed, all else equal. Statistical significance is assessed using the Wald test. Overall goodness-of-fit is evaluated through McFadden's pseudo-R^2 and the Akaike Information Criterion (AIC).

3.4 Machine Learning Algorithms

Three supervised algorithms are employed to exploit all 21 explanatory variables and compare their predictive performance in classifying employment status.

Random Forest

Introduced by Leo Breiman (2001), this algorithm relies on an ensemble of decision trees built on bootstrap sub-samples of the data (bagging), with random variable selection at each split. In our case, $B = 500$ trees are estimated, with a maximum depth of $d = 6$ and a minimum leaf size of 25 observations. Variable importance is assessed using the mean decrease in Gini impurity.

XGBoost

Developed by Tianqi Chen (2016), XGBoost (Extreme Gradient Boosting) is a sequential boosting method that builds trees iteratively, each new tree correcting the errors of the previous model. The model is estimated with 300 trees, a learning rate of $\eta = 0.06$, and a maximum depth of 4. L1 (Lasso) and L2 (Ridge) regularisation terms control model complexity and limit overfitting.

SVM (RBF Kernel)

Support Vector Machines maximise the margin between classes in a potentially high-dimensional representation space. Through the Gaussian (RBF) kernel, SVMs can model non-linear decision boundaries. The regularisation parameter $C = 1.0$ controls the trade-off between margin maximisation and classification error minimisation. The class_weight = balanced option corrects potential class imbalance by penalising errors on the minority class more heavily.

3.5 Model Evaluation

All models are evaluated using stratified 5-fold cross-validation on five metrics: accuracy, precision, recall, F1-score, and AUC-ROC. Stratified cross-validation ensures proportional representation of classes (unemployed/employed) in each fold.

The dataset D is partitioned into $K = 5$ subsets of approximately equal size, each fold maintaining the same proportion of unemployed individuals. At iteration k, the model is trained on $D \setminus Fk$ and evaluated on Fk. The final performance is the average across all K folds: $\bar{M} = (1/K)\ \Sigma\ M(k)$, where M(k) denotes the value of metric M at fold k.

JOURNAL OF SMART ECONOMIC GROWTH

www.jseg.ro ISSN: 2537-141X Volume 11, Number 2, Year 2026

Accuracy = (TP + TN) / (TP + TN + FP + FN). Precision = TP / (TP + FP). Recall = TP / (TP + FN). F1 = 2·(Precision × Recall) / (Precision + Recall). AUC-ROC is the area under the Receiver Operating Characteristic curve, equal to the probability that the model ranks a randomly chosen unemployed individual above a randomly chosen employed individual. A value of 0.5 corresponds to a random classifier, while 1.0 indicates perfect discrimination.

4. EMPIRICAL RESULTS

4.1 Descriptive Statistics

Unemployment Rates by Sub-Group

Table 2. Unemployment rates by sub-group (n = 1,200)

Variable	Category	N unemployed	Rate (%)
Gender	Female	216	**38.5**
	Male	155	24.3
Education level	Doctorate	16	**33.3**
	Bachelor's	205	32.8
	Master's	152	28.8
Field of study	Economics/Management	182	**34.5**
	Sciences	73	27.7
	Engineering	88	26.2
	Other	28	38.9
Internships	No	107	**30.7**
	Yes	264	31.0
Training adequacy	No	142	28.2
	Partial	127	33.1
	Yes	113	**36.2**

The bivariate analysis highlights several salient findings. First, the gender disparity is striking: female graduates exhibit an unemployment rate of 38.5%, 14.2 percentage points higher than men (24.3%). Second, bachelor's degree holders (32.8%) are slightly more exposed than master's degree holders (28.8%), confirming the persistence of the diploma paradox, albeit in a muted form with this large sample. Third, engineering graduates record the lowest unemployment rate (26.2%), compared with 34.5% for economics/management graduates (Mestiri, 2024).

Labour Market Perceptions

Table 3. Labour market perception scores

Question	% Disagree	% Neutral	% Agree	Mean
Market opportunities	58%	29%	12%	2.32
Employment matching degree	67%	24%	10%	2.12
Satisfactory wages	71%	20%	9%	2.05
Market fairness	68%	22%	10%	2.12
Growth → employment	20%	31%	50%	3.45
Lack of experience	5%	20%	75%	4.10
Recruitment transparency	53%	32%	16%	2.47
Personal connections	4%	15%	82%	4.25
University preparation	47%	33%	20%	2.63
Internships → insertion	12%	24%	64%	3.97

Respondents express widespread pessimism regarding market opportunities (mean = 2.32), the match between employment and degree (2.12), and recruitment transparency (2.47). In contrast, near unanimity (82%) acknowledges the predominant role of personal connections (,4.25), and 75% of respondents identify lack of experience as a major obstacle (4.10). The university is considered ill-prepared by 47% of respondents (2.63).

4.2 Logistic Regression Results

Table 4. Logistic regression results

Variable	$\hat{\beta}$	SE	OR	p-value	Sig.
Intercept	−1.659	0.601	0.190	0.006	***
Gender (Female = 1)	+0.657	0.129	**1.929**	<0.001	***
Age (25–30 years)	+0.089	0.148	1.093	0.551	ns
Age (> 30 years)	−0.194	0.186	0.824	0.296	ns
Master's level	−0.180	0.132	0.836	0.175	ns
Doctorate level	−0.022	0.320	0.979	0.946	ns

Engineering field	−0.344	0.156	**0.709**	0.028	**
Sciences field	−0.143	0.166	0.867	0.388	ns
Adequacy: No	−0.362	0.159	**0.696**	0.023	**
Adequacy: Partial	−0.278	0.165	0.757	0.093	*
No internship	+0.036	0.142	1.037	0.799	ns
Market opportunities	−0.074	0.063	0.929	0.240	ns
Employment matching degree	+0.065	0.066	1.067	0.325	ns
Growth/employment	+0.080	0.060	1.083	0.182	ns
Lack of experience	+0.062	0.071	1.064	0.380	ns
Transparency	−0.022	0.063	0.978	0.729	ns
Personal connections	+0.148	0.077	**1.159**	0.056	*
University preparation	−0.049	0.061	0.952	0.419	ns

**** $p<0.01$; ** $p<0.05$; * $p<0.10$; ns = not significant. McFadden pseudo-$R^2 = 0.035$; AIC = 1,468.6; χ^2LLR = 51.7, $p < 0.001$; N = 1,200*

The main econometric results highlight several significant determinants of graduate unemployment. First, female gender emerges as a disadvantageous factor: with an odds ratio (OR) of 1.929, a female graduate has, all else equal, a 92.9% higher risk of unemployment than a male graduate, validating hypothesis H2. Second, an engineering background exerts a protective effect: graduates in this field display a 29.1% lower risk of unemployment (OR = 0.709) compared with economics and management graduates. Third, education–employment mismatch (category "no") is associated with an OR of 0.696, indicating a 30.4% reduction in the risk of unemployment compared with individuals reporting adequacy between their training and the labour market; this result is explained by a selection effect, as graduates who consider their training adequate tend to hold higher salary expectations and voluntarily extend their job search period. Finally, the variable related to personal connections (Q15) shows a significant positive effect (OR = 1.159), suggesting that a stronger perception of the importance of social networks increases the unemployment risk by 15.9%, highlighting the potentially discriminating role of social capital in access to employment.

4.3 Machine Learning Model Performance

Table 5. Comparison of predictive performance (stratified 5-fold cross-validation)

Model	Accuracy	Precision	Recall	F1	AUC-ROC
Logistic regression	67.7%	25.1%	2.7%	4.7%	0.582
Probit model	67.7%	25.1%	2.7%	4.7%	0.582
Random Forest	69.1%	—	—	—	0.584
XGBoost	63.3%	33.5%	18.3%	**23.7%**	**0.556**
SVM (RBF)	58.0%	38.2%	56.1%	**45.3%**	**0.588**

Note: The balanced-weight SVM maximises recall (56.1%) and AUC (0.588).

A first observation stands out: the predictability of unemployment from the questionnaire variables alone remains moderate, with AUC-ROC values ranging from 0.556 to 0.588. This is unsurprising: it reflects the existence of unobserved determinants that necessarily escape any declarative survey, whether informal relational networks, personal recommendations, psychological factors, or local labour market conditions.

Beyond this general finding, the models differ markedly depending on the performance criterion considered. Logistic regression and the probit model produce strictly identical results (67.7% accuracy, AUC of 0.582), confirming their asymptotic equivalence on this type of data. Random Forest achieves the highest overall accuracy (69.1%), but this performance is misleading: its near-zero recall reveals a massive bias towards the majority class (the non-unemployed), at the expense of detecting individuals actually in unemployment.

It is the RBF-kernel SVM with balanced class weighting that offers the best trade-off between sensitivity and discriminating power, with a recall of 56.1% and an AUC of 0.588. Its ability to capture non-linear relationships in feature space allows it to better identify unemployed individuals where linear models fail. XGBoost occupies an intermediate position, with an F1 score of 23.7% reflecting a modest balance between precision and recall.

Ultimately, the choice of model depends on the analytical objective: if the goal is to maximise the detection of unemployed individuals — the central issue for targeted public policy — the

SVM is the most appropriate specification. If, on the other hand, econometric interpretability of coefficients is the priority, logistic regression remains the essential reference.

4.4 Variable Importance — Random Forest

Table 6. Variable importance — Random Forest (Gini impurity)

Rank	Variable	Gini	Interpretation
1	Gender (Female)	**0.231**	Dominant determinant
2	Q18 — University preparation	0.078	University readiness
3	Q11 — Growth/employment	0.072	Macroeconomic outlook
4	Q7 — Market opportunities	0.072	Perception of prospects
5	Q15 — Personal connections	0.069	Social capital
6	Q13 — Lack of experience	0.067	Entry barrier
7	Q14 — Transparency	0.061	Recruitment opacity
8	ADQ_NO — Mismatch	0.055	Training mismatch
9	Q8 — Employment by degree	0.052	Perceived downgrading
10	MASTER — Master's level	0.051	Human capital
11	AGE2 — 25–30 years	0.048	Cohort effect
12	ENGIN — Engineering	0.044	Sectoral advantage

Female gender dominates with a Gini importance of 0.231, more than double the second factor (Q18, 0.078). This result partially confirms H3: it is not personal connections but gender that emerges as the primary discriminating determinant in our large sample. Labour market perception variables (Q7, Q11, Q13, Q14, Q15) collectively account for a cumulative importance of 0.341, underscoring that the subjective dimensions of the labour market have discriminating power comparable to structural characteristics.

5. DISCUSSION AND POLICY IMPLICATIONS

The most salient result of our study is undoubtedly the predominance of gender as a determinant of unemployment, with an odds ratio of 1.929 in the logistic model and a Gini importance of 0.231 in the Random Forest. A Tunisian female graduate thus faces nearly twice the risk of unemployment as a comparable male graduate. In light of these inequalities, strong structural measures are required: differentiated fiscal incentives for firms hiring female graduates, strengthened legal mechanisms for recourse against discrimination, employer awareness campaigns targeting gender biases in recruitment, and the development of childcare provision and flexible working arrangements enabling women to reconcile professional and family life.

Education–employment mismatch appears significant across all estimated models. The apparently counter-intuitive result — an odds ratio below 1 for the group reporting no mismatch — is explained by a selection effect: graduates who consider their training adequate hold higher salary expectations and voluntarily extend their job search, waiting for a position commensurate with their qualifications. This voluntary waiting phenomenon, highlighted in Mortensen's (1994) matching models, suggests that programme reform will not suffice without a parallel reform of salary expectations and career guidance mechanisms.

Reliance on personal connections in job searching is significant at the 10% level in the logistic model and has a Gini importance of 0.069 in the Random Forest. This result confirms that social capital constitutes a determinant of unemployment in its own right, penalising candidates lacking networks. Institutional reforms are therefore required: generalisation of public calls for applications, creation of a national job vacancy platform based on transparency and merit, and strengthened sanctions against discriminatory recruitment practices.

Finally, the engineering field presents a significantly lower unemployment risk than average, with an odds ratio of 0.709. Orientation and information policies for secondary school students towards high-potential fields — engineering, computer science, green energy — combined with financial incentives such as excellence scholarships or apprenticeship contracts, would help rebalance the structure of training supply and better align academic pathways with the actual needs of the labour market.

6. CONCLUSION

This study has analysed the determinants of graduate unemployment in Tunisia by combining classical econometrics (logistic regression) with machine learning (Random Forest, XGBoost, SVM) on a sample of 1,200 graduates. Five main findings emerge.

First, the overall unemployment rate stands at 30.9%, but this average conceals a gender disparity of 14.2 percentage points: women exhibit a rate of 38.5% against 24.3% for men, revealing a deep structural inequality in the Tunisian labour market.

Second, econometric results confirm that female gender is the most powerful determinant of unemployment . Engineering emerges as a protective factor, as does the match between the training pursued and the employment sought , both variables significantly reducing the unemployment risk.

Third, the machine learning algorithms confirm and reinforce these econometric results: gender emerges as the absolutely dominant determinant, with a Gini importance of 0.231 in the Random Forest model.

Fourth, in terms of predictive performance, machine learning models achieve AUC-ROC values between 0.56 and 0.59. This moderate but above-chance level indicates that observed variables capture only part of reality, and that important unobserved determinants such as social networks, informal discrimination, and local labour market conditions play a significant role in graduates' occupational integration.

Fifth, four priority reform directions emerge: the implementation of active policies against gender discrimination in hiring; the professionalisation of higher education to strengthen education–employment adequacy; the improvement of recruitment process transparency; and the promotion of fields with strong employment potential.

Limitations and Future Research

Although the sample of 1,200 graduates provides adequate statistical power, the convenience sampling method may introduce selection biases that limit the generalisability of the findings. Future research should draw on comprehensive administrative data and incorporate a longitudinal dimension to track integration trajectories over time. The use of causal identification methods, such as instrumental variables or difference-in-differences, would moreover be indispensable for establishing robust causal relationships and moving beyond the level of observed correlations.

CONFLICTS OF INTEREST AND PLAGIARISM: The authors declare no conflict of interest and plagiarism.

. REFERENCES

Alrabba, M. I. M. (2017). The determinants of unemployment rate in Jordan: a multivariate approach. International Journal of Economics and Finance, 9(11), 109–117.

Arrow, K. J. (1973). Higher education as a filter. Journal of Public Economics, 2(3), 193–216.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis. Columbia University Press.

International Labour Organization (ILO). (2022). Youth Employment Report for North Africa. Geneva: ILO.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), Handbook of Theory and Research for the Sociology of Education (pp. 241–258). Greenwood.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Chalfin, A., Hansen, B., Lerner, J., & Parker, L. (2016). Productivity and selection of human capital with machine learning. American Economic Review, 106(5), 124–127.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.

Granovetter, M. S. (1973). The strength of weak ties. American Journal of Sociology, 78(6), 1360–1380.

National Institute of Statistics (INS). (2023). National Employment Survey. Tunis.

Mestiri, S. (2024). How to Use Stata Software. University of Monastir Press.

Mestiri, S. (2025a). A comparative analysis of SVM and neural networks for credit risk prediction. SSRN Electronic Journal.

Mestiri, S. (2025b). Modeling the unemployment rate using the panel ARDL. Journal of Smart Economic Growth, 10(1), 67–77.

Mincer, J. (1974). Schooling, Experience and Earnings. National Bureau of Economic Research.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. Review of Economic Studies, 61(3), 397–415.

OECD. (2023). Employment Outlook: Tunisia. OECD Publishing.

Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3), 355–374.

World Bank. (2020). Tunisia Economic Monitor. Washington, D.C.: World Bank.

World Bank. (2022). Tunisia Labor Market Report. Washington, D.C.: World Bank.