

А.А. Бова,
кандидат соціологічних наук,
старший науковий співробітник

ВИЗНАЧЕННЯ ФАКТОРІВ ЗАДОВОЛЕНОСТІ ПРАЦІВНИКІВ СИСТЕМИ МВС УКРАЇНИ СЛУЖБОЮ ЗА ДОПОМОГОЮ МЕТОДІВ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ⁹

У статті наводяться результати логістичної регресії, методу групового врахування аргументів, алгоритму індукції правил та алгоритму дерев класифікації QUEST, які пояснюють рівень задоволеності працівників системи органів МВС України. Як емпіричну базу використано дані опитування, проведеного у п'яти регіонах України у жовтні 2015 року. Встановлено, що багатовимірні статистичні моделі мають приблизно однакову середню точність (близько 70 % об'єктів вибірки класифіковано правильно). Більшим лаконізмом вирізняються моделі класифікаційних правил та методу групового врахування аргументів. Кожна зі статистичних моделей визначила власний набір кращих пояснюючих змінних. Спільним для всіх моделей є виокремлення стабільної роботи та фактору задоволеності службою.

Ключові слова: задоволеність службою, поліція, логістична регресія, метод групового врахування аргументів, індукція правил, дерева класифікації.

В статті приводяться результати логістической регресии, метода группового учета аргументов, алгоритма индукции правил и алгоритма деревьев классификации QUEST, которые объясняют уровень удовлетворенности работников системы органов МВД Украины. В качестве эмпирической базы использованы данные опроса общественного мнения, проведенный в пяти регионах Украины в октябре 2015 года. Установлено, что многомерные статистические модели имеют примерно одинаковую среднюю точность (около 70 % объектов выборки классифицированы правильно). Большим лаконизмом отличаются модели классификационных правил и метода группового учета аргументов. Каждая из статистических моделей определила собственный набор лучших объясняющих переменных.

Ключевые слова: удовлетворенность службой, полиция, логистическая регрессия, метод группового учета аргументов, индукция правил, деревья классификации.

Paper presents the results of logistic regression, the method of group account of arguments, induction of algorithm rules and classification tree algorithm QUEST, which explain the level of satisfaction of the Ukrainian Interior Ministry employees. As an empirical basis data from a public opinion poll, conducted in five regions of Ukraine in October 2015 is used. It was found that the multivariate statistical models have approximately the same high accuracy (about 70 % of the sample objects are classified correctly). More laconic classification model are classification tree and the Group method of data handling. Each of statistical models defines its own set of the best

⁹ Продовження. Початок у попередньому номері.

explanatory variables. Common to all of them is the release of the stable operation of a service satisfaction factor.

Keywords: *satisfaction with service, police, logistic regression, group method of data handling, rules induction, classifications trees.*

Визначення задоволеності працівників роботою є важливим завданням менеджменту, адже результати анонімних соціологічних опитувань дають змогу чіткіше визначити мотиви діяльності, різноманітні конфлікти, що існують в середині організації, емоційні та когнітивні реакції на оплату праці та преміювання, правила, стилі управління тощо. Вивчення факторів задоволеності поліцейських службою здійснювалося у різних країнах. Науковцями розроблено низку теорій, типологій, шкал та статистичних моделей, що фіксують задоволеність працівників організацій. Так, зокрема, задоволеність службою серед поліцейських аналізувалася в контексті сприйняття ефективності діяльності, проявів лідерських якостей та стилів керівництва, психологічного стресу, виміру впливу на загальний рівень задоволеності різноманітних особистісних факторів та факторів середовища.

При статистичній обробці даних соціологічних опитувань у випадку порядкової залежної змінної найчастіше використовується однофакторний дисперсійний аналіз та регресійний аналіз, а у випадку категоріальної змінної – аналіз таблиць сполучень та логістична регресія. Додамо, що незалежні змінні, що характеризують діяльність, узагальнювалися в інтегральні індекси у тому числі за допомогою факторного аналізу.

У багатовимірному статистичному аналізі даних розроблено значну кількість різноманітних класифікаційних моделей, які, зокрема, розрізняються за характером отримання оцінок для вибірки алгоритмів розпізнавання образів – глобальні (наприклад, дискримінантний аналіз), глобально-локальні (дерева класифікацій) та локальні (k -найближчих сусідів), та можливостями змістовної інтерпретації. Низка методів класифікації дозволяє робити прогнози для нових даних лише за наявності генерації коду на мовах програмування.

Сучасні програми машинного навчання дають змогу зіставляти точність та складність різних класифікаційних моделей, формувати колективи моделей та застосовувати мета-алгоритми (бегінг та бустінг). У соціальних науках “прозорість” результатів моделювання досягається за рахунок аналізу рівнянь (класифікаційних функцій) або логічних правил. Такі властивості мають, зокрема, методи: наївний байєвський класифікатор; множинний номінальний аналіз; дискримінантний аналіз; логістична регресія; дерева класифікацій; індукція логічних правил; дискримінантний аналіз – аналіз відповідностей; дискримінантний аналіз – проекція на латентні структури.

Переваги глобальних моделей полягають в отриманні рівняння для всієї вибірки та його застосування для емпіричних даних, зібраних іншими дослідниками, наприклад, крос-національних досліджень, оцінювати статистичну значимість коефіцієнтів. Глобально-локальні моделі дають змогу побудувати різноманітні евристичні типології, що складаються з однорідних множин, в яких максимізується значення категорій залежної змінної. Дерева рішень, маючи переваги у наочному представленні результатів класифікації, є вкрай нестійкими при зміні її обсягу. Попри наявність різноманітних алгоритмів класифікації точність моделей щодо соціологічних даних є невеликою в силу недосконалості теорії (яка гуртується у тому числі і на узагальненні емпіричних даних), особливостей методик та неоднорідності вибірки. Результати роботи сучасних (непрозорих) алгоритмів

машинного навчання, що орієнтовані на обробку великих обсягів інформації, можуть перевершувати за точністю передбачення теоретичні конструкції [3].

У жовтні 2015 року на виконання доручення Міністра внутрішніх справ України здійснено, за участю фахівців Державного науково-дослідного інституту МВС України, вивчення громадської думки працівників міліції в ГУМВС України у Харківській, Одеській, Дніпропетровській, Львівській областях та ГУМВС м. Києва [3]. Метою роботи є побудова та змістовна інтерпретація класифікаційних моделей, що пояснюють задоволеність працівників системи органів МВС України службою. При цьому використано різні форми представлення результатів статистичного моделювання (рівняння логістичної регресії, класифікаційна функція, знайдена за допомогою методу групового урахування аргументів, набір логічних закономірностей, дерево класифікації) з автоматичною селекцією найбільш інформативних змінних. Багатовимірна статистична обробка даних здійснювалася у пакетах SPSS, GMDH Shell та WEKA.

Методичний інструментарій складено на підставі пропозицій працівників міліції та досвіду відповідних опитувань поліції Польщі. Анкета містила низку запитань, які фіксували загальний рівень задоволеності та задоволеність різними сторонами життєдіяльності атестованих працівників органів системи МВС України.

Зокрема, запитання *“Якою мірою Ви задоволені своєю роботою у цілому?”* включало п'ять градацій (*“цілком задоволений”, “скоріше задоволений”, “скоріше не задоволений”, “цілком не задоволений”* та *“важко відповісти”*). Шкала була перетворена на дихотомічну (не задоволений – 0, задоволений – 1) шляхом об'єднання відповідних рівнів та видалення із масиву сукупності респондентів, котрі обрали альтернативу *“важко відповісти”*. Таким чином частка задоволених респондентів роботою (службою) у цілому складає 33,8 %, а не задоволених – 66,2 % ($n = 1154$).

Респондентам було запропоновано відповісти також на п'ять запитань із сумісними альтернативами, які характеризують різні аспекти задоволеності/незадоволеності службою, особливості діяльності та недоліки забезпечення, а саме: (i) *“Які з аспектів найбільше впливають на Вашу задоволеність?”* (18 альтернатив); (ii) *“Які з аспектів найбільше впливають на Вашу незадоволеність?”* (18 альтернатив); (iii) *“Які види забезпечення Вашої службової діяльності найбільшою мірою потребують поліпшення?”* (14 альтернатив); (iv) *“Що Вам найбільше заважає у щоденній роботі/службі?”* (16 альтернатив); (v) *“Що найбільше впливає на кар'єрне зростання у Вашому підрозділі?”* (9 альтернатив).

Методика соціологічного опитування містила схожі за змістом запитання аби виявити найбільш чутливі формулювання, що найкращою мірою співвідносяться із загальним рівнем задоволеності роботою працівників МВС України. Для цього здійснено кореляційний аналіз 66 дихотомічних змінних (не включалися альтернативи *“інше”, “немає таких аспектів”, “нічого не заважає”*) зі змінною, яка відбиває задоволеність службою. Виходячи зі значень парних коефіцієнтів кореляції (статистично значущі на рівні статистичної значущості $\alpha = 0,01$), найбільшою мірою негативно на рівень задоволеності роботою впливає вибір таких альтернатив: *“родинні зв'язки з керівництвом або іншими впливовими особами”, “відсутність робочої стабільності – постійні організаційні зміни”* (iv) – $r = -0,2$ та *“відсутність можливості кар'єрного зростання”* – $r = -0,19$. Позитивний вплив здійснює погодження з твердженнями щодо *“стабільної роботи”* (i) – $r = 0,20$, *“можливостей професійної кар'єри”* (ii) – $r = 0,17$, *“гарних стосунків з керівництвом та колегами по роботі”* (iii) – $r = 0,16$. Можна припустити, що ядро задоволеності/незадоволеності

службою складають фактори, що відбивають стабільність роботи та просування кар'єрними сходами.

Зазначимо також, що за безпосередніми висловлюваннями респондентів основними факторами задоволення службою є стабільна робота (56 %), гарні стосунки з керівництвом та колегами по роботі (32,3 %), можливість професійної кар'єри (28,2 %) та професійного самовдосконалення (25,6 %), а також велика відпустка (24 %), а незадоволення – відносно низька заробітна плата (82,1 %), неналежне технічне забезпечення (53,6 %), значна кількість документації (42,8 %).

За результатами соціологічного опитування вплив на кар'єрне зростання у підрозділі родинних зв'язків з керівництвом або іншими впливовими особами відмітило 15,4 % опитаних. Нестабільність роботи порушує ритм, знижує професійну мотивацію, а непрозорість при призначеннях на посади порушує принципи конкурентної боротьби. Своєю чергою це може вплинути на підвищення психологічного напруження та міжособистісні конфлікти через різний обсяг виконуваної роботи працівниками, які обіймають однакові посади, ускладнення контролю за дисципліною близьких до керівництва осіб тощо.

Кореляція соціально-демографічних ознак (стать, вік, стаж роботи) із загальним рівнем задоволеності виявилася низькою та статистично-незначущою. Водночас, з підвищенням матеріального статусу, зростає і задоволеність службою ($r = -0,15$, $\alpha = 0,01$). Існують також деякі відмінності у задоволенні службою (роботою) працівників різних служб (коефіцієнт кореляції V Крамера = 0,11, $\alpha = 0,05$).

Гнучкі налаштування сучасних алгоритмів аналізу даних уможливило побудову кількох альтернативних багатовимірних статистичних моделей з відбором системи найбільш інформативних змінних. Для виявлення загальних факторів, що впливають на задоволеність роботою в цілому, до бінарної покрокової логістичної регресії включено 25 незалежних (пояснюючих) змінних (ні – 0, так – 1), значення коефіцієнту кореляції яких із залежною змінною є більшим за модулем, ніж 0,11. Такий спосіб відбору змінних до моделі додатково гарантує стійкість рішення.

Показники якості моделі R^2 Кокса і Снелла = 0,21, R^2 Нейджелкерка = 0,30. Загалом вірно класифіковано 71,9 % об'єктів вибірки, зокрема серед категорії не задоволених службою – 70,3 % та категорії задоволених службою – 72,8 %. Як бачимо з таблиці 1 до остаточного рівняння увійшли 14 змінних (статистично значущі за критерієм Вальда), які характеризують різноманітні фактори задоволеності службою, зокрема явища, що заважають у повсякденній роботі та механізми кар'єрного зростання, а брак певних видів забезпечення не має впливу на загальний рівень задоволеності.

Таблиця 1

Результати логістичної регресії, що пояснює задоволеність службою працівників системи МВС України

Змінна \ Статистика	B	Стандартна похибка	Критерій Вальда	Ступені свободи	Рівень значущості	Exp(B)
родинні зв'язки з керівництвом (v)	-0,59	0,21	8,31	1	0,004	0,55
невпевненість у роботі (ii)	-0,79	0,21	13,75	1	0,000	0,45
відсутність робочої стабільності (iv)	-0,58	0,16	14,06	1	0,000	0,56
стабільна робота (i)	0,71	0,15	23,78	1	0,000	2,03
понаднормова робота, відсутність вихідних (iv)	-0,59	0,15	15,55	1	0,000	0,55

Змінна \ Статистика	B	Стандартна похибка	Критерій Вальда	Ступені свободи	Значущість	Exp(B)
можливість професійної кар'єри (<i>i</i>)	0,50	0,18	7,83	1	0,005	1,66
гарні стосунки з керівництвом та колегами по роботі (<i>i</i>)	0,74	0,17	19,30	1	0,000	2,10
доступ до Інтернету (<i>iii</i>)	-0,32	0,16	3,95	1	0,047	0,73
необхідність скласти велику кількість документів (<i>iv</i>)	-0,50	0,16	10,42	1	0,001	0,60
погана система перепідготовки, навчання, тренінгів (<i>ii</i>)	-0,75	0,23	10,75	1	0,001	0,47
можливість професійного самовдосконалення (<i>i</i>)	0,56	0,18	9,31	1	0,002	1,76
відданість службі (<i>v</i>)	0,44	0,16	7,56	1	0,006	1,55
погані стосунки з керівництвом (<i>ii</i>)	-0,96	0,39	6,17	1	0,013	0,38
необхідність дотримуватися суворої дисципліни (<i>iv</i>)	-1,22	0,46	7,03	1	0,008	0,30
константа	0,77	0,14	29,28	1	0,000	2,16

Враховуючи значення коефіцієнтів логістичної регресії та відповідних відношень шансів, можна зробити висновок, що найбільшу вагу у позитивному впливові на задоволеність мають такі змінні, як стабільна робота, гарні стосунки з керівництвом та колегами по роботі, можливість професійного самовдосконалення, можливість професійної кар'єри, відданість службі як головна умова кар'єрного зростання. Натомість зменшує задоволеність персоналу вибір альтернатив “необхідність дотримуватися суворої дисципліни”, “погані стосунки з керівництвом”, “невпевненість у роботі”, “погана система перепідготовки, навчання, тренінгів”.

Надалі до статистичної обробки включалися усі 66 незалежних змінних. За допомогою методу групового урахування аргументів (швидкий комбінаторний алгоритм) побудовано класифікаційну модель за умови збалансування класів залежної змінної (формула 1). Вибірку випадковим чином розділено на навчальну (80 % спостережень) та екзаменаційну (20 %). На навчальній вибірці 68 % випадків вірно класифіковано за моделлю, у тому числі для класу задоволених роботою – 72 %, незадоволених роботою – 62 %. На екзаменаційній вибірці відповідно вірно класифіковано 71 %, та відповідно за класами – 75 % та 60 % респондентів.

$$\hat{Y} = 0,6 + 0,14X_1 + 0,07X_2 - 0,18X_3 - 0,16X_4 + 0,02X_5 - 0,1X_6 - 0,14X_7 - 0,22X_8 - 0,06X_9 - 0,1X_{10}, \quad (1)$$

де:

- Y – задоволеність службою;
- X_1 – стабільна робота (*i*);
- X_2 – суспільне значення роботи (*i*);
- X_3 – непевненість у роботі (*ii*);
- X_4 – відсутність можливості кар'єрного зростання (*ii*);
- X_5 – забезпечення транспортом (*iii*);

- $X6$ – необхідність складати велику кількість документів (iv);
- $X7$ – відсутність робочої стабільності – постійні організаційні зміни (iv);
- $X8$ – необхідність дотримуватися суворої дисципліни (iv);
- $X9$ – незрозумілі службові вказівки (iv);
- $X10$ – робота “на статистику” (iv).

Зазначимо, що модель, отримана за допомогою методу групового урахування аргументів, включає лише 10 предикторів (близько трьох змінних коефіцієнт має дуже низьке значення, що фактично дорівнює нулю). Оскільки усі незалежні змінні виміряні у дихотомічній шкалі, беручи за основу значення коефіцієнтів, найбільше підвищує рівень задоволеності стабільна робота, а знижує – необхідність дотримуватися суворої дисципліни, невпевненість у роботі, відсутність можливості кар’єрного зростання.

Крім рівняння, сформованого для даних загалом, доцільно виявити закономірності, що відносяться до окремих підмножин вибірки. Алгоритм індукції правил *JRip* виявив 4 логічних умови, які сформовані комбінаціями альтернатив 4 змінних, а саме:

1. Якщо невпевненість у роботі (ii) = 1 та відсутність робочої стабільності – постійні організаційні зміни (iv) = 1 та відданість службі (v) = 0, то задоволеність роботи в цілому = 0.

2. Якщо родинні зв’язки з керівництвом або іншими впливовими особами (v) = 1 та відсутність робочої стабільності – постійні організаційні зміни (iv) = 1, то задоволеність роботи в цілому = 0.

3. Якщо невпевненість у роботі (ii) = 1, то задоволеність роботи в цілому = 0.

4. Інакше задоволеність роботи в цілому = 1.

На тренувальному наборі даних алгоритмом вірно розпізнано 72,6 % об’єктів, у тому числі задоволених службою – 80,8 %, незадоволених – 52,0 %. Алгоритм виявив докладні правила класифікації респондентів, незадоволених службою, виходячи з вибору ними тих чи інших альтернатив. Так, зокрема, серед незадоволених службою в абсолютному вираженні виявилось у правилі 1 та у правилі 3, де опитані респонденти відмітили як фактор незадоволеності невпевненість у роботі та відсутність робочої стабільності.

Оскільки кількість задоволених респондентів перевищує майже у два рази кількість незадоволених, для підвищення точності передбачення кількості незадоволених респондентів, в алгоритмі *QUEST* була задана однакова апріорна ймовірність належності до класів. На навчальній вибірці правильно класифіковано 70,7 % об’єктів вибірки, у тому числі задоволених – 78,5 % та не задоволених – 55,4 %.

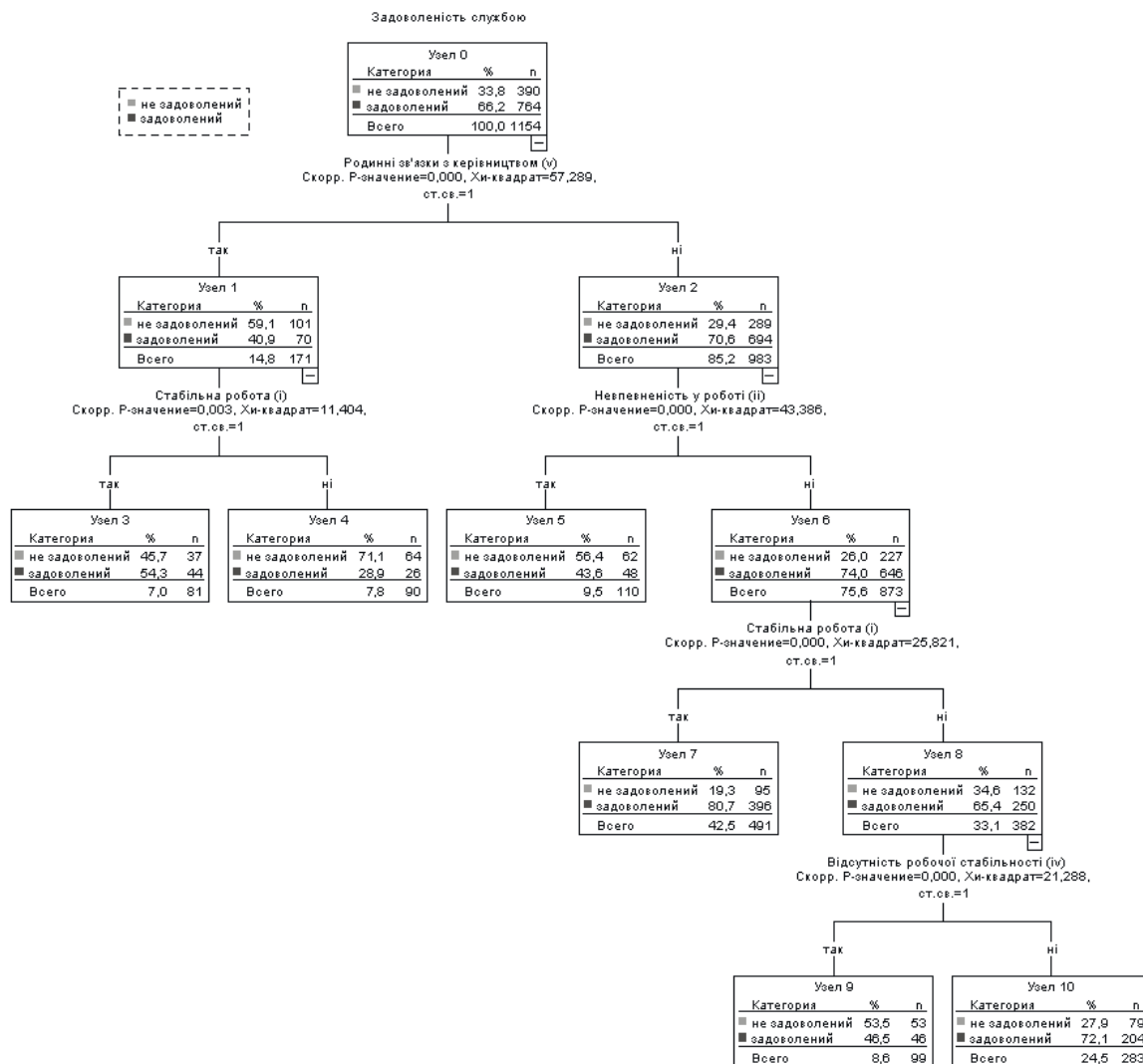


Рис. 1. Дерево класифікації QUEST, що пояснює задоволеність/незадоволеність працівників системи МВС України службою, %

Дерево класифікації включає чотири незалежних змінних, завдяки значенням яких утворюється шість логічних правил, що розмежовують два класи (задоволених та незадоволених службою). Ці змінні стосуються загалом стабільності роботи та просування кар'єрними сходами. Відсутність стабільності, використання родинних зв'язків для здобуття більш високих посад може суттєвим чином порушувати принципи соціальної справедливості та вносити дисбаланс у роботу колективу, що збільшує частку незадоволених службою.

На підставі проведення статистичного аналізу встановлено, що велика кількість потенційних пояснюючих змінних дає можливість побудувати кілька моделей, які розрізняються за загальною точністю та коректністю класифікації класів, а також набором інформативних ознак. Класифікаційні моделі, сформовані у вигляді дерев рішень або логічних правил, наочніше відображають характер зв'язку між змінними. Можна стверджувати, що стабільність роботи та дотримання керівництвом принципів соціальної справедливості при призначенні працівників на посади та стабільність роботи збільшують у колективах правоохоронців частку задоволених службою. Виявлення системи факторів у різних сферах, що характеризують службову діяльність та впливають на загальний рівень задоволеності, дає змогу повнішою мірою врахувати нагальні потреби персоналу організації.

Знайдені моделі мають приблизно однакову середню точність (близько 70 % об'єктів вибірки класифіковано правильно). Більшим лаконізмом вирізняються дерева класифікації, індукція правил та метод групового урахування аргументів. Способами підвищення стійкості рішення є побудова рівняння на змінних, що мають найбільше значення коефіцієнту кореляції з цільовою змінною, його тестування на різних підвбірках. Кожна зі статистичних моделей визначила власний набір кращих пояснюючих змінних. Спільним для всіх моделей є вибір альтернатив, що відбивають стабільність роботи як один із головних факторів задоволеності службою.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Chan J., Bennet-Moses L. Is Big Data challenging criminology? // *Theoretical Criminology* (Published online). – 2015. – 19 May.– 19 p.

Отримано 13.10.2016