

УДК 004.67: 664.69

¹ М.М. Буряченко

студент

² С.В. Грибков

к.т.н., доцент, доцент кафедри інформаційних систем

^{1,2} Національний університет харчових технологій, Київ

ВИКОРИСТАННЯ СЕРЕДОВИЩА R ДЛЯ АНАЛІЗУ ТА ВИЯВЛЕННЯ ПРИЧИН ВИНИКНЕННЯ НЕКОНДИЦІЙНОЇ ПРОДУКЦІЇ У МАКАРОННОМУ ВИРОБНИЦТВІ

Вступ. Сьогодні макаронне виробництво в Україні проходить новий виток розвитку та адаптації до сучасних економічних умов ринку, що диктується економічним станом в країні та вимогами до високих європейських стандартів якості. Українські виробники конкурують не тільки між собою, ай зовнішніми постачальниками макаронних виробів. Необхідно відмітити, що макаронне виробництво характеризується великими обсягами та кожне окреме підприємство є унікальним, адже відрізняється виробничо-технічним устаткуванням, що має дуже широкий спектр характеристик. Устаткування різниться по технологіям виготовлення макаронних виробів, видам продукції (короткорізнана, довгорізнана та штамповочна) та потужності, що знаходиться в межах 100 кг до 10 тон на годину. Головною тенденцією українських виробників для відстоювання позицій ринку є підвищення якості кінцевого продукту, тому актуальною задачею є поглиблений контроль якості макаронних виробів шляхом аналізу виробничо-технологічного циклу для усунення усіх негативних факторів впливу та підвищення техніко-економічних показників виробництва.

Результати та обговорення. Авторами в роботі здійснено акцентування на аналізі виробничого циклу та вплив його характеристик на кінцевий продукт, але крім цього враховуються показники сировини, не зважаючи на те, що будь-яка сировина прийнята на виробництво, повинна відповідати усім вимогам якості. Крім цього необхідно відмітити, що макаронне виробництво відноситься до потокового не перервного виробництва, адже весь технологічний цикл відбувається таким чином, що його не можна зупиняти, а протікання технологічних процесів чітко розбито на певні етапи, на яких є можливість впливати на технологічний процес.

Враховуючи економічний стан, авторами для проведення інтелектуального аналізу даних була обрана мова аналізу даних R, що є вільним аналогом середовища S-PLUS та є лідером для статичному аналізу останнім часом, адже має широке коло застосування при розв'язку різних статистико-аналітичних задач пов'язаних з проведенням прогнозування, визначення ризику та ймовірності, створення рекомендацій, пошук послідовностей, групування та інше. Мова R дозволяє провести усі основні етапи процесу побудови моделі інтелектуального аналізу даних, що є частиною більш масштабного процесу, в який входять всі завдання, від формулювання питань щодо даних і створення моделі для відповідей

на ці питання до розгортання моделі в робочому середовищі. Основними базовими етапами проведення інтелектуального аналізу даних є: постановка задачі, підготовка даних, перегляд даних, побудова моделей, дослідження і перевірка моделей, розгортання і оновлення моделей.

Для проведення аналіз була обрано метод кластеризації, що підтримується мовою R та реалізовано пакетами. Завдання кластеризації - полягає в розподілі безлічі об'єктів на групи (кластери) схожих за параметрами, що можуть заздалегідь невідомі і визначатися в ході побудови кластерів виходячи зі ступеня близькості об'єднуються об'єктів за сукупністю параметрів.

Вхідними значеннями для масивів обробки даних використано звіти по виробництву за зміну та результати лабораторних досліджень якості готової продукції та знаті напівфабрикати підчас технологічного процесу.

При використанні засобів мови даних R для аналізу обробки вхідних масивів даних має наступну послідовність дій: обробка даних; кластеризація даних, що полягає у виділенні груп однорідних об'єктів в залежності від обраного методу та пакету застосування; оцінки якості результатів кластеризації де відбувається оцінка щільності і локальності отриманих кластерів; представлення результатів у візуальній формі різноманітних типів графіків з використанням пакету «cluster»; інтерпретація результатів.

В процесі дослідження було випробувано наступні методи: ієрархічна кластеризація та метод k-середніх з пакету «cluster»; метод bootstrap-реплікації на базі пакету «rvclust»; дискримінантний аналіз з пакету «lda»; дерева рішень з пакету «rpart», метод Random Forest з пакету «randomForest». За результатами дослідження не має можливості надати перевагу одному з методів, тому що тільки комплексне застосування забезпечує отримання більш точного результату, що дає змогу здійснювати об'єктивні та зважені висновки по кожній ситуації.

Висновки і перспективи подальших досліджень.

Для отримання всебічного розуміння ситуації буває не достатньо використати один метод при аналізі виробничих ситуацій для виявлення причин отримання некондиційної продукції, тому пропонується використовувати декілька комбінацій різних методів, що дасть глибше проаналізувати кожну окрему ситуацію. Недоліком середовища R є обмеження обсягу даних до декількох сотень тисяч записів, оскільки, обчислення зберігаються в оперативній пам'яті, але попередня обробка вхідних даних її усуває. В подальшому планується продовжити роботи в даному напрямку та розширити застосування середовища та мови R для аналізу виробничих ситуацій для підприємств макаронної галузі

Список використаних джерел

1. Шипунов А.Б., Балдін Е.М. Аналіз даних з R [Текст] / А. Б. Шипунов. – М.: ДМК Пресс, 2008. – 60 с.
2. Gan, G. Data Clustering: Theory, Algorithms, and Applications [Text] / G. Gan, C. Ma, J. Wu – Philadelphia: SIAM, 2007. – 466 p.