

УДК 004.42

О.Қ. Хўжаев, Ш.Н. Саидрасулов, О.Т.Алламов

МАЪЛУМОТЛАРНИ ИНТЕЛЛЕКТУАЛ ТАҲЛИЛИ ДАСТУРИЙ ВОСИТАЛАРИНИНГ ҚЎЛЛАНИШ СОҲАЛАРИ

Мазкур ишда ҳозирда Ўзбекистон Республикасида дастурий воситалар соҳасида қилинаётган ишлар, ишлатилаётган дастурий воситалар ва уларни маълумотларни интеллектуал таҳлили билан боғлиқ жиҳатлари берилган. Маълумотларни интеллектуал таҳлили дастурий воситалари ва уларнинг қўлланилиш соҳалари ҳақида солиштирма таҳлил асосида маълумотлар келтирилган. Ҳозирда мавжуд маълумотларни интеллектуал таҳлили дастурий таъминотлари умумий вазибалари ва имкониятлари келтириб ўтилган. Ундан ташқари энг кўп ишлатиладиган маълумотларни интеллектуал таҳлили дастурий воситаларни имкониятлари бўйича қиссий таҳлили келтирилган.

Калит сўзлар: Маълумотларни интеллектуал таҳлили, KNIME, Rapid miner, Orange, Tanagra, синфларга ажратиш, башоратлаш, синфлаштириш.

Республикамизда ахборот коммуникация технологияларини ривожлантириш ва тадбиқ этиш учун бир нечта қонунчилик базаси ишлаб чиқилган:

- Ахборотлаштириш тўғрисида;
- Электрон рақамли имзо тўғрисида;
- Электрон ҳужжат айланиши тўғрисида;
- Электрон тижорат тўғрисида;
- Электрон тўловлар тўғрисида;

Бу ишлаб чиқарилган қонун ва қарорларни самараси ўлароқ жамиятнинг турли соҳаларига тегишли бўлган маълумотларини йиғиш, марказлашган ҳолда уларга ишлов бериш, уларни бошқариш каби ишларни амалга оширувчи дастурий таъминотлар сони ва улардан фойдаланиш кундан-кунга ошиб бормоқда:

- Корхона, ташкилотларда электрон ҳужжат алмашилини таъминловчи дастурий таъминотлар- E-hujjat, Docs Open, Microsoft SharePoint Portal Server Optima, Workflow, "БОСС-Референт", "Евфрат", DocFlow (NCI Project), NauDoc Lotus Domino.doc, Globus Professional ва бошқалар.

- Таълим соҳаси учун яратилган виртуал таълим тизимлари- Chamilo, Claroline, Dokeos, Moodle, ILIAS ва бошқалар.

- Соғлиқни сақлаш муассасалари учун ишлаб чиқилган дастурий таъминотлар- OpenEMR, GNU health, FreeMED, Hospital OS, HRHIS ва бошқалар.

Бундай дастурий воситаларни яна бошқа соҳа корхоналари учун ҳам кўплаб келтириш мумкин ва уларнинг кўпчилиги очик кодли GNU/GPL лицензияси асосида тарқатиладиган дастурий таъминотлар ҳисобланиб булардан бемалол фойдаланиш мумкин.

Бу каби ишлар фақатгина маълумотлар базасини йиғиш ва марказлашган маълумотлар базасини яратган ҳолда шу бўйича турли сўровларга жавоб бериш, ҳар хил форматдаги ҳисоботлар яратиш имконини яратади холос. Булардан ташқари биз йиғилган маълумотлар базасини таҳлил қилиб корхона ёки ташкилот ривожини башорат қилишимиз ва маълумотларни таҳлил қилиш мобайнида маълумотларни синфлаштириш ва синфларга ажратишимиз мумкин. Бу масалалар маълумотларни интеллектуал таҳлилининг асосий масалалари саналади. Ҳозирда маълумотларни интеллектуал таҳлили усул ва алгоритмларини қуйидаги соҳа масалаларига қўллаш яхши самара беради:

1) *Банк соҳаси:* банк соҳасидаги катта ҳажмдаги маълумотлар базасидан фойдаланиб келгусидаги жараёнларни башорат қилиш мумкин. Кредит муаммоларини ечишда ҳам қўллаш мумкин. Фойдаланувчилар пластик карталарларидаги маблағлар сарфини ҳам башорат қилишда муҳим ҳисобланади.

2) *Саноат ва ишлаб чиқариш соҳаси:* мавсумий хомашё келадиган корхоналар учун олдинги йиллардаги кундалик хомашё келиши ҳақидаги маълумотларга асосланган ҳолда жорий йил учун ишчи кучини тақсимлаш. Ишлаб чиқариш корхоналари учун талаблар базаси асосида бугун ва эрта учун қанча маҳсулот ишлаб чиқарса мақсадга мувофиқлигини башорат қилиш муҳим аҳамият касб этади.

3) *Маркетинг соҳаси:* Супермаркетлар учун миқдорларни демографик классификациясини аниқлаш, ҳар бир маҳсулотга бўлган келгусидаги талабни айтиш бизнес ривожини учун зарур ҳисобланади.

4) *Соғлиқни сақлаш соҳаси:* Беморларни демографик классификация қилиш, касалликларнинг симптомлари бўйича бир-бирига боғлиқлик даражаларини ўрганиш, олдиндан йиғилган маълумотлар базаси асосида шифохоналарга беморларни келишини башорат қилиш ва шу асосида дори воситалари тақсимотини йўлга қўйиш.

5) *Суғурта соҳаси:* Бу соҳада ҳам маълумотларни интеллектуал таҳлилининг классификация ва башоратлаш масалаларини қўллаш яхши самара беради, суғурта полиси олувчи миқдорларни синфлаштириш ва суғуртага бўлган талабни мавсум бўйича башоратлаш.

6) *Қонунчилик соҳаси:* Жиноятлар ҳақидаги маълумотлар базасини таҳлил қилган ҳолда регионлар бўйича жиноятларни синфлаштириш, жиноятчиликнинг энг асосий атрибутларини аниқлаш каби муаммоларни ечишда ёрдам беради.

7) *Мудофаа соҳаси:* маълумотларнинг интеллектуал таҳлили армия ва уни ҳарбий техникасининг бир жойдан бошқа жойга кўчириш ва моддий ресурслар тақсимотини башоратлашда қўлланилади.

8) *Брокерлик ташкилотлари ва қимматбаҳо қозғалар бозори:* бу соҳада маълумотларнинг интеллектуал таҳлили акциялар бозорида нархларнинг қимматлаши ёки арзонлашишини башорат қилиши ва шунга асосланиб сотиш ёки сотиб олишни маслаҳат берувчи эксперт системаларни яратишда қўлланилади.

9) *Интерактив хизматларни яратиш соҳасида*: фойдаланувчилар учун турли мақсадларда маълумотларни интеллектуал таҳлил қилишда, башоратлашда ва синфларга ажратишда қўллаш яхши самара беради. Масалан, транспорт воситалари ҳаракатини тартибга солиш ва шаҳарларда тирбандликларни камайтириш учун фойдаланувчиларга бир жойдан бошқасига борадиган энг мақбул маршрутларни танлашда муҳим ҳисобланади.

Маълумотларни интеллектуал таҳлили - оддий маълумот анализи воситаларидан то мураккаб, ҳақиқий модели ва катта маълумотлар орасидаги боғланишларни аниқлайдиган воситаларни ўз ичига олади. Бундай воситаларга статистик моделлар, математик алгоритм ва машинали ўқитиш алгоритмларни қўшиш мумкин. Ўз навбатида, маълумотларни интеллектуал таҳлили жуда кўп маълумотлар тўпламини бошқариш, таҳлил қилиш ва башоратлашларни ҳам ўз ичига олади[1].

Маълумотларни интеллектуал таҳлили масалаларини ечиш учун мўлжалланган дастурий таъминотларнинг тарихи қисқа, лекин шу қисқа давр мобайнида кўпгина ишлар нашр қилинган. Маълумотларни интеллектуал таҳлили атамасининг ўзи 90-йиллар ўрталарида пайдо бўлди. Дастлаб яратилган маълумотларни интеллектуал таҳлили дастурларга DOS ёки UNIX тизимида ишловчи, маълумотларни текст файлдан ўқиб C4.5 алгоритми бўйича классификация масаласини ечадиган дастурни келтириш мумкин[2].

Ҳозирги кунга келиб маълумотларни интеллектуал таҳлили масалалари ечиш учун барча яратилган дастурий таъминотлар қуйдагича умумий вазибаларга эга:

- Маълумотларни дастлабки кўриб чиқиш учун статистик таҳлил воситалари;
- Маълумотларни визуализация қилиш учун турли усуллар мавжуд:
 - Гистограммалар;
 - Диаграммалар;
 - Параллел координаталар системаси;
 - Графикларни тақсимлаш ва бошқалар;
- Маълумотларга ишлов бериш учун стандарт компоненталар (маълумотлар базаси билан уланиш ва унга турли хил сўровлар бериш, дискретлаш ва тартиблаш функциялари, қисм тўпламларни танлаш ва бошқалар);
- Маълумотларни ўқув танланмаларсиз таҳлили методлари (асосий компонентли таҳлил, турли кластеризация алгоритмлари, ассоциатив қоидаларни чиқариш, қисм гуруҳлар таҳлили методлари ва бошқалар);
- Маълумотларни ўқув танланмали таҳлили методлари (классификация қоидалари ва дараклар, таянч векторлар тизими, нави Байес классификаторлари ва бошқалар);
- Моделни баҳолаш воситалари (классификация аниқлиги, сезувчанлик, сонли таҳлил натижалари ва бошқалар)

- Бошқариладиган ва бошқарилмайдиган таҳлил асосида ишланган маълумотларни башоратлаш моделлари;

- Моделни қандайдир стандарт форматда сақлаш (масалан PMML формати);

- Натижалар бўйича турли хил ҳисоботларни чиқариш.

Юқорида санаб ўтилган модулларни баъзиларини ўзида жамлаган дастурий воситалар мавжуд. Қуйида кўп фойдаланувчилар ишлатадиган маълумотларни интеллектуал таҳлили дастурий воситалари ва уларни имкониятлари бўйича маълумотларни таққослаб келтириб ўтамыз.

1) **KNIME** – маълумотларни интеллектуал таҳлили масалалари учун очик кодли дастурий восита ҳисобланади. Java дастурлаш тили асосида қурилган. Асосий афзалликлари маълумотлар таҳлилини лойиҳалаш жуда қулайлиги ва кўпгина маълумотларни интеллектуал таҳлилининг машхур усуллари ва алгоритмларини ўзида жамлаганлиги билан ажралиб туради. Ундан ташқари R[4], Weka каби бошқа маълумотларни интеллектуал таҳлил дастурлари билан ҳам интеграциялаша олади[1][2].

2) **Rapid Miner** – маълумотларни интеллектуал таҳлили масалалари учун очик кодли дастурий восита ҳисобланади. У операцион тизим танламайдиган ва унинг махсулотлари билан интеграциялаша оладиган яхлит тизим бўлиб, яхши график интерфейсга эга. Маълумотлар таҳлилидаги жараёнлар модулларга ажратилган ҳолда, график шаклда тасвирларни беради.

3) **Weka** – маълумотларни интеллектуал таҳлили масалаларини ечиш учун бир нечта машинали ўқитиш алгоритмларини ўзида жамлаган дастурий восита ҳисобланади[5]. **Weka** Java дастурлаш тили асосида қурилган дастурий восита бўлиб, турли операцион тизимларда ҳам ишлай олади ва маълумотларга дастлабки ишлов бериш, классификация, кластеризация, регрессия, ассоциатив хотира ва визуализация воситаларидан иборат[6].

4) **Tanagra** – маълумотларни интеллектуал таҳлил қила оладиган машинали ўқитиш, статистик ўқитишнинг бир нечта усул ва алгоритмларини ўзида жамлаган очик кодли дастурий восита ҳисобланади. У асосан илмий изланиш олиб борадиганлар учун ишлаб чиқилган. Унинг юқорида келтирилган дастурлардан устунлик қилувчи имкониятларидан бири бу фойдаланувчилар ҳам ўзлари яратган методларни тизимга киритишлари мумкин ва ўзи яратган усул натижаларини бошқа алгоритмлар ёрдамида олинган натижалар билан солиштиришлари мумкин.

5) **Orange** – очик кодли дастурий воситаси ҳисобланиб, турли хил операцион тизимларда ишлай олади. Унда маълумотлар таҳлилини лойиҳалаш яхши йўлга қўйилган, визуал дастурлаш ва Python скриптларини қўллаб қувватлайди. Бу дастур асосан турли хил визуализация услублари борлиги билан ҳам машхур ҳисобланади[7]. Келтирилган дастурлар солиштирма таҳлилини 1-жадвалда келтирамыз.

Маълумотларни интеллектуал таҳлил дастурий воситаларининг солиштирма таҳлили

Жараён	KNIME	Rapid Miner	Weka	Tanagra	Orange
Маълумотлар базасини ўқитилувчи ва тестловчи тўпламларга ажратиш	Чекланган имкониятлар асосида ажрата олади	Чекланган имкониятлар асосида ажрата олади	Чекланган имкониятлар асосида ажрата олади	Чекланган имкониятлар асосида ажрата олади	Чекланган имкониятлар асосида ажрата олади
Масштабланиш дескрипторлари	Бундай имконият мавжуд	Бундай имконият мавжуд	Бундай имконият мавжуд эмас	Бундай имконият мавжуд эмас	Бундай имконият мавжуд эмас
Танлаш дескрипторлари	Бундай методлар йўқ	Бундай имконият мавжуд	Бундай имконият мавжуд	Фақат логистик регрессия учун имкониятлар мавжуд.	Бундай методлар йўқ
Статистик/Эвристик методларнинг оптимизациялаш параметри	Бундай имконият мавжуд эмас	Бундай имконият мавжуд	Бундай имконият мавжуд эмас	Бундай имконият мавжуд эмас	Бундай имконият мавжуд эмас
Моделни текширишда cross-validation методдан фойдалангани ёки мустақил тестлаш тўпланидан?	Фақат хатоликлар ўзгаришининг чекланган методлари мавжуд	Бундай имконият мавжуд	Бу имконият мавжуд созланган теслаш усулини сақлаш имкони йўқ ва буни ҳар бир маълумотлар базаси учун қайта қилиш зарур	Мустақил тестлаш тўпламлари ёрдамида текшириш имконияти йўқ	Бу имконият мавжуд созланган теслаш усулини сақлаш имкони йўқ ва буни ҳар бир маълумотлар базаси учун қайта қилиш зарур

Юқорида таксослаб ўтган маълумотларнинг интелектуал таҳлили дастурий воситалари текин дастурий таъминотлар ҳисобланади. Уларнинг имкониятлари бўйича солиштиришда кўринадики Rapid Miner дастури бошқа дастурларга қараганда мукамалроқ ҳисобланади.

Таҳлиллар натижалари шуни кўрсатадики корхона ёки ташкилотларда иш жараёнини баҳолаш учун ҳам соҳа корхоналари учун яратилган дастурий таъминотлардан ташқари шу дастурлар асосида яратилган маълумотлар базаларини интеллектуал таҳлил дастурий воситаларидан фойдаланиш ёки мавжуд соҳа корхоналари дастурий воситаларига маълумотларнинг интеллектуал таҳлили модулларини қўшган ҳолда яратиш муҳим ҳисобланади.

Фойдаланилган адабиётлар

1. Blaz Zupan, Janez Demsar. Open-source tools for data mining. Clin lab Med (2008)
2. Hoffman P.E., Grinstein G.G., Marx K.E., DNA visual and analytic datamining. Phoenix (AZ):In Proc. IEEE Visualization; 1997.
3. Witten IH, Frank E. Datamining: practical machine learning tools and techniques with Java implementations. 2nd edition. San Francisco (CA): Morgan Kaufmann; 2005.
4. <http://www.r-project.org/> accessed on February 2013.
5. <http://www.cs.waikato.ac.nz/ml/weka/> accessed on February 2013.
6. [http://en.wikipedia.org/wiki/Weka_\(machine_learning\)](http://en.wikipedia.org/wiki/Weka_(machine_learning)) accessed on February 2013.
7. <http://orange.biolab.si/features/> accessed on February 2013.
8. <http://rapid-i.com/content/view/181/190/> accessed on February 2013.
9. <http://eric.univ-yon2.fr/~ricco/tanagra/en/tanagra.html> accessed on February 2013.
10. <http://blog.revolutionanalytics.com/2012/05/r-tops-data-mining-poll.html> accessed on February 2013.
11. <http://www.kdnuggets.com/2012/05/top-analytics-data-mining-big-data-software.html> accessed on February 2013.

12. <http://invenis.net/blog/industry-news/data-mining-future-trends-predicted-2012> accessed on February 2013.

Хўжаев Отабек Қадамбоевич

Мустақил изланувчи, “Ахборот технологияларининг дастурий таъминоти” кафедраси, Мухаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети (ТАТУ)
Тел.: (+99894)-658-47-71;
Эл. почта: otabek.hujaev@gmail.com;

Саидрасулов Шерзод Норбой ўғли

Мустақил изланувчи, “Ахборот технологияларининг дастурий таъминоти” кафедраси, Мухаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети (ТАТУ)
Тел.: (+99899)-8338363,
Эл. почта: sh.saidrasulov@mitc.uz

Алламов Ойбек Турабаевич

Мустақил изланувчи, “Ахборот технологияларининг дастурий таъминоти” кафедраси, Мухаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети (ТАТУ)
Тел.: (+99893)-758-84-8);
Эл. почта: oybek.allamov@gmail.com;

Khodayev O.K , Saidrasulov Sh.N., Allamov O.T. Implementing Of Software For Intellectual Data Analysis And Their Comparative Analysis

In this paper provides information on activities in the field of software development in the Republic of Uzbekistan, about the used software tools and their relationship to data mining. Provides detailed information about the software associated with data mining. Given the common tasks in the application of existing software data mining. Besides the comparative analysis of the most popular data mining tools.

Keywords: data mining, KNIME, Rapid Miner, Orange, Tanagra, classification, forecasting, clustering, methods.