

Data mining технологи ашиглан их дээд сургуулийн сургалтын албаны үйл ажиллагааг боловсронгуй болгох асуудалд

Мянгадын Тунгалаг
Мэдээллийн Технологийн салбар
Мэдээлэл, Холбоо Технологийн Сургууль, ШУТИС
Улаанбаатар, Монгол
И-мэйл: tungalag@must.edu.mn

Хураангуй - Их дээд сургуулийн сургалтын албаны асар их хэмжээний мэдээлэлд боловсруулалт хийж, оновчтой шийдвэр гаргалтанд туслах сургалтын албаны ухаалаг систем болон түүнд шаардлагатай Data mining технологийн талаар судлан үзлээ. Судалгааны явцад Data mining технологи нь байгууллагын өгөгдлийн сан болон өгөгдлийн агуулах дээр суурилан хүн бүрт хэрэгтэй мэдээллийг гарган, гаргасан мэдээллээ төрөл бүрийн төхөөрөмжүүд дээр хүлээн авах боломжтой нь ажиглагдсан. Энэ үндсэн дээр Data mining ашиглан их дээд сургуулийн сургалтын албаны үйл ажиллагааг боловсронгуй болгох талаар өгүүлэв.

Түлхүүр үг - Data mining, өгөгдөл, өгөгдлийн сан, шийдвэрийн мод

I. ОРШИЛ

Дэлхий нийтийн чиг хандлагаар боловсролын салбар тэр дундаа их дээд сургуулиуд мэдээлэл, холбооны технологийн хөгжлийг амжилттай ашиглан асар их мэдээллээс хэрэгтэйг нь авч үлдэн тэдгээр мэдээллүүд дээрээ тулгуурлан ирээдүйн дүр төрх, стратеги боловсруулахад шаардлагатай мэдээллийг олж авч байна. Гадаадын төрийн болон төрийн бус их дээд сургуулиудын хүрээ томорч, цаашилбал глобал шинж чанартай болж байгаа ба түүний гол хөдөлгөгч хүч нь өрсөлдөөн юм. Энэ өрсөлдөөнд давуу чанараа бий болгох, үйл ажиллагаагаа сайжруулах, шинэчлэх ажлуудыг мэдээллийн технологи хангаж байдаг. Манай улсын хувьд их дээд сургуулиуд сургалтын албаны үйл ажиллагаагаа боловсронгуй болгох шинэ арга замыг олж хэрэгжүүлэх нь нэн даруй шийдэх асуудлуудын нэг юм. Үүнтэй холбоотойгоор өгөгдлийн сангуудыг нэг буюу хэд хэдэн баазад төвлөрүүлж сүлжээ болон онлайн хэлбэрээр хандан өгөгдөл боловсруулах шинэ технологи гарч ирж байна. Өгөгдлийн сангийн удирлагын системд дэлхийн том том програм хангамжийн компаниуд өгөгдлийн тандалтыг өргөнөөр ашиглаж байна.

II. СУДАЛГААНЫ АЖЛЫН АРГА ЗҮЙ

Өгөгдлийн тандалт технологийг дараах шинжлэх ухааны судалгааны аргууд дээр тулгуурлан хийдэг. Үүнд:

- Ажиглалтын арга (Мэдээлэл цуглуулах болон нөөцлөх)
- Таамаглалын арга (Таамаглал дэвшүүлэх)
- Туршилтын арга (Цуглуулсан өгөгдөлд туршилт хийх)
- Анализ, синтезийн арга

Бидний хувьд ажиглалтын болон туршилтын аргыг түлхүү хэрэглэсэн.

III. ӨГӨГДЛИЙН ТАНДАЛТ

Өгөгдлийн тандалт бол шинэ ойлголт төдийгүй компьютерийн шинжлэх ухааны салбарын нэг хэсэг юм. Өгөгдлийн тандалт (*Data Mining*) нь их хэмжээний өгөгдлүүдээс өгөгдлийн сангийн систем, математик, статистикийн арга, загвар таних аргуудаар мэдлэг олж авах аргачлал юм. Өөрөөр хэлбэл өгөгдлийн сангаас хамгийн хэрэгцээтэй мэдлэг цуглуулах арга юм. Маш их хэмжээний өгөгдлүүдээс чухал загваруудыг олж авахын тулд автомат болон хагас автомат аргаар судалгаа шинжилгээ хийж болдог. Статистик болон хиймэл оюуныг өгөгдлийн сангийн удирдлагын системтэй хослуулсан шинж чанартай, их хэмжээний өгөгдлүүдээс загвар гарган авах үйл явцыг өгөгдлийн тандалт гэнэ. Үүнийг бас өгөгдлийн сангаас мэдлэг илрүүлэх процесс гэж нэрлэдэг.

Өгөгдлийн тандалт гэдэг нь боловсруулаагүй өгөгдөл дотроос далд зүй тогтол, утгын уялдаа холбоог нээн илрүүлэх процесс ба статистик анализ, машины сургалтын арга зэргийг багтаадаг хэдий ч өгөгдлийн автомат анализ илүүтэй ашигладаг.

Өгөгдлийн тандалт нь компьютер техник хэрэгсэл ашиглан автоматаар шинжилгээ хийж өгөгдлийн санд агуулагдаж байгаа их хэмжээний өгөгдлүүдээс мэдлэг гарган авах процесс юм.

Өгөгдлийн тандалт нь маш өргөн хүрээг хамарч шинжилгээ судалгаа хийх тул мэдлэг илрүүлэх гэсэн дан ганц ойлголттой бас андуурч болохгүй. Зарчмын

хувьд өгөгдлийн сангаас мэдлэг илрүүлэх нь өгөгдлийн тандалтын шинжлэх ухааны аргуудын нэг юм.

Өгөгдлийг ерөнхийд нь дахин ашиглах зорилгоор хадгалсаар ирсэн. Өгөгдлийн тандалт нь шинэ ойлголт боловч технологи нь бол тийм биш. Өгөгдлөөс загвар гарган авах нь хэдэн зууны туршид гараар хийгдэж байсан. Жишээ нь Бейсийн теором (1700-аад оноос), регрессийн шинжилгээ (1800-аад оноос) зэрэг нь өгөгдлийн загварыг тодорхойлох эртний аргууд билээ. Компьютер үүсч мэдээллийн технологи эрчимтэй хөгжсөнөөр өгөгдөл боловсруулалтанд нейрон сүлжээ, бүлэглэлт, генийн алгоритм (1950-аад оноос), шийдвэрийн мод (1960-аад оноос) болон тулгуурт векторын машин (1990-ээд оноос) зэрэг ойлголтууд шинээр нээлт болон гарч ирсэн.

Өгөгдлийн тандалт нь 1980-аад оны сүүлийн үед үүссэн. 1990 оны сүүл үед өгөгдлийн тандалтыг ашиглан гайхалтай алхмуудыг хийсэн бөгөөд одоо өгөгдлийн тандалтыг компьютерийн шинжлэх ухаанаар дамжуулан шинжлэх ухаан, эрүүл мэнд, санхүү, бизнесийн ухаан, эдийн засаг, биоинформатик, хүн судлал, сансар судлал зэрэгт өргөн ашиглаж байна.

IV. ӨГӨГДЛИЙН ТАНДАЛТ ХИЙХ АРГУУД

Өгөгдлийн тандалтын модуль нь өгөгдлүүдийн ерөнхий шинжийг тодорхойлох, өгөгдлүүдийн холбоо, хамааралд дүн шинжилгээ хийх, өгөгдлүүдийг нийтлэг шинжүүдээр нь бүлэг болгон ангилах, тодорхой үзүүлэлтүүдийн тусламжтай урьдчилан таамаглах, өгөгдлүүдийг нийтлэг шинжүүрээр нь бүлэглэх, хэт ялгаатай өгөгдлүүдийг ялгах үүрэг гүйцэтгэнэ.

Өгөгдлийн тандалтыг хяналттай ба хяналтгүй хоёр үндсэн төрөлд хуваадаг. Хяналттай аргачлал нь зорилго нь тодорхой, судлаач нь зорилгоо биелүүлдэг байх ёстой. Хяналтгүй аргачлал нь тодорхой зорилгогүй, судлаач нь тэрхүү боловсруулаагүй өгөгдлүүдээс мэдлэгийг олж авах хэрэгтэй болдог.

Өгөгдлийн тандалтын аргуудыг ямар үүрэг гүйцэтгэж байгаагаар нь хоёр ангилдаг.

А. Таамаглах аргууд (Predictive tasks)-
Өгөгдлүүдийн тусламжтай урьдчилсан дүгнэлт гаргах

Б. Тодорхойлох аргууд (Descriptive tasks)-
Өгөгдлийн сан дахь өгөгдлүүдийн ерөнхий шинж чанарыг тодорхойлох

1. Ангилал (Classification)

Ангилал нь нөхцөл бүрээр ангилан тодорхойлох үйлдлийг хэлнэ. Нөхцөл бүр нь атрибутуудын олонлогоос бүрдэх бөгөөд эдгээрийн нэг нь класс атрибут байдаг. Энэ арга нь оролтын атрибутуудын функцээр класс атрибутыг тодорхойлон загвар гарган авахыг шаарддаг. Мөн өгөгдлийн тандалтын хяналттай алгоритмуудыг ашигладаг. Шийдвэрийн мод, нейрон сүлжээ зэрэг нь ангиллын алгоритмууд юм.

2. Бүлэглэлт (Clustering)

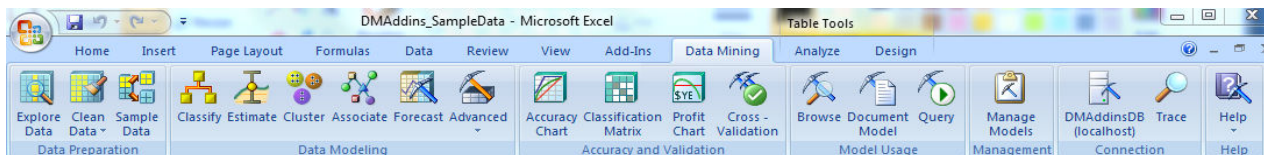
Ямар нэг тодорхой, тогтоосон, төсөөтэй шинжүүрээр (хэмжээ) объектыг бүлэглэхийг объектын бүлэглэлт (cluster) гэдэг. Ингэж бүлэглэх процессыг ангилал хийх (clustering) буюу хуваалт гэж нэрлэдэг. Тухайн бүлэглэлтийг цаашид мөн объект мэт авч үзэж болно. Өөрөөр хэлбэл бүлэглэлт хийх аргын үүрэг нь өгөгдөл ба объектын өгөгдсөн олонлогоос бүлэг буюу ангийг (дэд олонлог, бүлэг, анги хэсэг) гаргаж авах явдал юм.

V. MICROSOFT SQL SERVER 2010 DATA MINING ADD-INS ПРОГРАМ

Microsoft корпорацийн програм хангамж хөгжүүлэгчид MSSQL Server програмын Analysis services бүлэгт өгөгдлийн тандалт технологийг бүрэн хэмжээгээр програмчилж чаджээ. MSSQL Server дэх өгөгдлийн тандалт нь хэрэглэгчдэд иж бүрэн, хамгийн сайн боловруулалтыг хийх боломжийг Office 2010 хэрэглээний програмаар дамжуулан олгож байна. Энэ програмаа Microsoft SQL 2010 Data Mining Add-ins гэж нэрлэсэн байна.

MICROSOFT OFFICE БА ӨГӨГДЛИЙН ТАНДАЛТ

Office 2010 дээр Data Mining, Analyze цэсүүдийг нэмж оруулсан



Зураг 1. Data Mining цэс

Уг цэс нь 6 хэсэгтэй. Хэсэг тус бүр нь өөрийн гэсэн өгөгдлийн тандалт хийх шинжилгээний аргууд бүхий хэрэгслүүдээс бүрдэнэ.

1) Өгөгдлийн урьдчилсан боловсруулалт (Data Preparation) хэсэг

2) Өгөгдлийг илрүүлэх хэрэгсэл (Explore data)

Өгөгдлийг илрүүлэх болон бэлдэх нь өгөгдлийн тандалтын чухал процесс юм. Өгөгдлийг нээж бэлдэх нь шийдвэрлэх гэж байгаа асуудлын таамаглалын эхлэл гэж болно. Статистикийн тархалт ашиглан өгөгдлийг дискретчлэн хуваадаг. Энэ нь хэт их мэдээллийг бүлэглэж өгөх юм.

3) Өгөгдлийг цэвэрлэх хэрэгсэл (Clean data)

Цэвэрлэх, дахин сэргээх, дискретчилэх үйлдлүүдийг хийх бөгөөд доорх 2 хэрэгслээс бүрдэнэ.

Үүнд:

- Дагалдах хэсгийг устгах хэрэгсэл (Remove Outliers)-хэвийн бус өгөгдөл нь өгөгдлийн тандалтын үр дүнгийн чанарт муугаар нөлөөлдөг. Бүрэн бус мэдээллийг өөрчлөх, тодорхойлох, утгаар нь нийлбэрчлэх, хуурмаг мэдээллийг устгах замаар өгөгдлийг цэвэрлэдэг.
- Relabelхэрэгсэл - Өгөгдлийгбүлэглэх, илэрхийлэх аргыг өөрчлөн анализ хийхэд дөхөм болгоход энэ хэрэгсэл туслана.

4) Өгөгдлийг турших хэрэгсэл (Sample data)

Шинэ өгөгдлийн олонлогоо хуваах замаар сургалтын болон тестийн өгөгдөл болгон өгөгдлийг боловсруулахад бэлэн болгодог

5) Өгөгдлийн загварчилгаа (Data Modeling) хэсэг

Стандарт өгөгдлийн тандалтын алгоритмуудын дагуу өгөгдлийг загварчлах үйлдлийг өгөгдлийн загварчилгааны хэсэг дэх хэрэгслүүдээр хийж гүйцэтгэдэг.

6) Ангилал хэрэгсэл (Classification)

Ангилал нь нөхцөл бүрээр ангилан тодорхойлох үйлдлийг гүйцэтгэдэг.Нөхцөл бүр нь атрибутуудын олонлогоос бүрдэх бөгөөд эдгээрийн нэг нь класс атрибут байдаг. Энэ арга нь оролтын атрибутуудын функцээр класс атрибутыг тодорхойлон загвар гарган авахыг шаарддаг.

7) Үнэлэлт хэрэгсэл (Estimate)

Тоон утга, цаг, огноогоор урьдчилан тооцоолох замаар загвар үүсгэдэг.

8) Бүлэглэлт хэрэгсэл (Cluster)

Ижил төстэй шинж чанартай мөрүүдийг бүлэглэх замаар загвар үүсгэдэг.

9) Харилцан хамаарал хэрэгсэл (Associate)

Хэд хэдэн харилцан үйлчлэлд ордог зүйлсийн хоорондын харилцан хамаарлыг тогтоож загвар үүсгэнэ.

10) Урьдчилан таамаглах хэрэгсэл (Forecast)

Цаг хугацааны үзүүлэлтийг таамаглаж /прогноз/ урьдчилан тооцоолох боломжтой график үүсгэдэг.

11) Гүнзгий түвшний анализ хэрэгсэл (Advanced)

Одоо байгаа өгөгдлийн бүтцийг ашиглан шинэ загвар үүсгэдэг. Энэ загвар нь нэг загварыг ашиглан хэд хэдэн янзын аргаар анализ хийх боломжийг олгодог.

12) Нарийвчлал ба Баталгаа (Accuracy and Validation) хэсэг

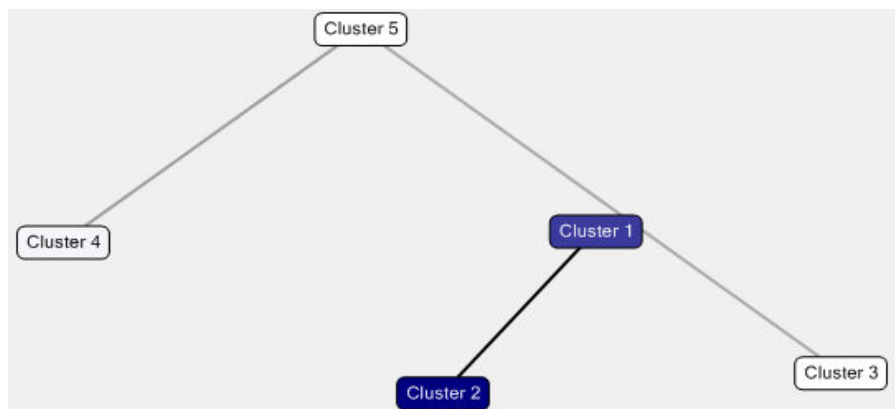
Уг хэсэг нь загвар болон өгөгдлийн чухал хандлагыг төрөл бүрийн график байдлаар харуулдаг. Ямар төрлийн графикаар харуулах нь өгөгдөл болон өгөгдөлдөө хийх өгөгдлийн тандалтын анализийн аргаас шалтгаална.

13) Нарийвчлалын график хэрэгсэл (Accuracy Chart)

Өгөгдлийн тандалтын шинж чанарыг өсөлтийн график болон тархалтын графикуудаар үнэлдэг.

VI. ӨГӨГДЛИЙН ТАНДАЛТЫН ТУРШИЛТ

ШУТИС-ийн ЭХС-ийн Эмнэлгийн тоног төхөөрөмжийн ашиглалт, Цахилгаан электроник болон Компьютерийн техникийн мэргэжлийн ангид сурч буй бакалавр оюутнуудын мэдээллийг сургалтын албанаас авч хэд хэдэн үзүүлэлтээр бүлэглэлт (Cluster), ангилал (Classify) хийж үзүүлээ (2-р Зураг).



Зураг 2. Бүлэглэлт диаграм

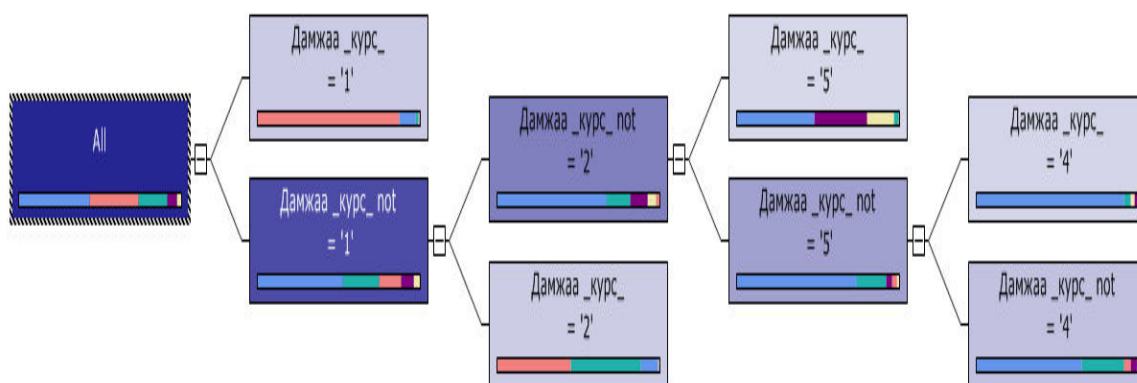
ХҮСНЭГТ 1. ОЮУТНЫ МЭДЭЭЛЛИЙГ ЗАРИМ ҮЗҮҮЛЭЛТ

Variables	States	Population (All)	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Дамжаа _курс_	3	126	48 %	0 %	47 %	2 %	87 %
Дамжаа _курс_	1	98	0 %	51 %	0 %	51 %	0 %
Дамжаа _курс_	2	93	1 %	49 %	0 %	47 %	0 %
Дамжаа _курс_	4	71	29 %	0 %	26 %	0 %	13 %
Дамжаа _курс_	5	60	22 %	0 %	27 %	0 %	0 %
Дамжаа _курс_	Missing	1	1 %	0 %	0 %	0 %	0 %
Мэргэжил	1	194	71 %	62 %	0 %	0 %	0 %
Мэргэжил	14	157	0 %	0 %	100 %	99 %	100 %
Мэргэжил	36	98	30 %	39 %	0 %	1 %	0 %
Мэргэжлийн чиглэл	Компьютерийн техник	194	71 %	62 %	0 %	0 %	0 %
Мэргэжлийн чиглэл	Цахилгаан электроник	157	0 %	0 %	100 %	99 %	100 %
Мэргэжлийн чиглэл	ЭТТА	98	30 %	39 %	0 %	1 %	0 %
Мэргэжлийн чиглэл2	Компьютерийн техник	194	71 %	62 %	0 %	0 %	0 %
Мэргэжлийн чиглэл2	Цахилгаан электроник	157	0 %	0 %	100 %	99 %	100 %
Мэргэжлийн чиглэл2	ЭТТА	98	30 %	39 %	0 %	1 %	0 %
Нас	Mean	21	21.81	19.23	21.97	18.96	20.27
Нас	Deviation	1.99	1.59	1.09	1.87	1.31	0.72
Салбар	52	255	30 %	39 %	100 %	100 %	100 %
Салбар	48	194	71 %	62 %	0 %	0 %	0 %
Суралцагчийн голч дүн	Mean	1.93	2.01	1.87	1.99	1.83	0.94
Суралцагчийн голч дүн	Deviation	1.13	1.1	1.03	1.26	1.16	1.28
Сургалтын төлбөрийн хэмжээ _мян_төг_	Mean	987,047.00	973,620.00	1,005,150.00	973,669.06	1,005,181.59	973,637.71
Сургалтын төлбөрийн хэмжээ _мян_төг_	Deviation	15,603.10					
Сургалтын төлбөрийн хэмжээ _мян_төг_3	Mean	987,047.00	973,620.00	1,005,150.00	973,669.06	1,005,181.59	973,637.71
Сургалтын төлбөрийн хэмжээ _мян_төг_3	Deviation	15,603.10					
Сургуулийн нэр	ШУТИС ЭХИС	449	100 %	100 %	100 %	100 %	100 %
Төрөлжсөн мэргэжил	0	292	100 %	100 %	0 %	1 %	0 %
Төрөлжсөн мэргэжил	1	157	0 %	0 %	100 %	99 %	100 %
Хүйс	Эр	312	62 %	62 %	83 %	85 %	74 %
Хүйс	Эм	130	36 %	37 %	15 %	15 %	26 %
Хүйс	Эр	5	2 %	2 %	0 %	0 %	0 %
Хүйс	Эм	1	1 %	0 %	0 %	0 %	0 %
Хүйс	Missing	1	0 %	0 %	1 %	0 %	0 %

ХҮСНЭГТ 2. КОМПЬЮТЕРИЙН ТЕХНИК БОЛОН ЭМНЭЛГИЙН ТОНОГ ТӨХӨӨРӨМЖИЙН АШИГЛАЛТ МЭРГЭЖЛИЙН ОЮУТНУУДЫН МЭДЭЭЛЛИЙН ЕРӨНХИЙ ҮЗҮҮЛЭЛТҮҮД

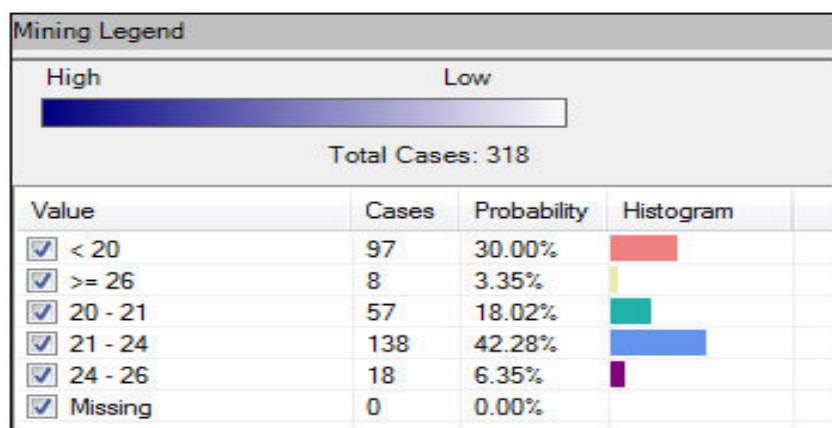
Variables	Values	Probability
Сургуулийн нэр	ШУТИС ЭХИС	100 %
Төрөлжсөн мэргэжил	0	100 %
Боловсролын зэрэг	D	100 %
Мэргэжлийн чиглэл	Компьютерийнтехник	70 %
Салбар	48	70 %
Мэргэжил	1	70 %
Хүйс	Эр	62 %
Сургалтын төлбөрийн хэмжээ мян төг 3	973,620 - 976,522	50 %
Сургалтын төлбөрийн хэмжээ мян төг	973,620 - 976,522	50 %
Аймаг нийслэл	Улаанбаатар	48 %
Дамжаа курс	3	48 %
Нас	23 – 26	45 %
Хүйс	Эм	36 %
Нас	21 – 22	31 %
Мэргэжил	36	30 %
Мэргэжлийн чиглэл	ЭГТА	30 %
Салбар	52	30 %
Дамжаа курс	4	29 %
Суралцагчийн голч дүн	1.9 - 2.7	26 %
Суралцагчийн голч дүн	1.2 - 1.9	25 %
Суралцагчийн голч дүн	2.7 - 4.0	24 %
Дамжаа курс	5	22 %
Суралцагчийн голч дүн	0.0 - 1.2	19 %
Нас	20	18 %

Шийдвэрийн мод алгоритмыг ашиглан нийт оюутны мэдээллийг курсээр нь ялган харуулав.



Зураг 3. Курсээр нь харуулсан байдал

Насны ангиллыг курсээр нь гаргасан загварыг Зураг 4-т харуулав.



Зураг 4. Шийдвэрийн мод алгоритм ашиглан насны ангилалыг
курсээр нь гаргасан загвар

ДҮГНЭЛТ

Их дээд сургуулиуд өөрсдийн үйл ажиллагаанд шаардлагатай мэдээллийг гаргаж авах түүнчлэн сургалтын бодлого, стратегийг тодорхойлох зорилгоор ашиглах системийг гаргаж авна. Энэ системийг боловсруулахад Data mining технологийг ашиглах боломжтой.

Аль ч их дээд сургуулиуд өөрсдийн сургалтын албаны үйл ажиллагааны чиглэлээс хамааран модулуудаас сонгон нэвтрүүлэх боломж өргөн байна.

Шаардлагатай мэдээллийг хэрхэн боловсруулах, ангилах, бүлэглэх алгоритмуудыг ашиглан сургалтын албаны мэдээллийн нэгдсэн үзүүлэлт гаргалаа.

АШИГЛАСАН МАТЕРИАЛ, НОМ ЗҮЙ

- [1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (references)
- [2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [4] K. Elissa, "Title of paper if known," unpublished.
- [5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989

- [1] Data Analysis Using SQL and Excel, Gordon S. Linoff, ISBN: 978-0-470-09951-3, USA
- [2] Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, May 12, 2005, China
- [3] Data Mining: Concepts and Techniques, 2nd ed. The

Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, March 2006.

ISBN

1-55860-901-6

1. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, University of Illinois at Urbana-Champaign, San Francisco, 2006
2. Roiger, Geatz, Data mining A Tutorial-Based Primer, 2003, page 5
3. Ж. Дашдорж, Ц. Мөнхсайхан, "Өгөгдлийн тандалтын (Data Mining) ангиллын онолын судалгаа", "Эрчим хүчний хөгжилд бидний оролцоо" бакалавр, магистр, доктор оюутны ЭШБХ, УБ хот, 2010 он, хуудас 215-220
4. Өгөгдлийн сан удирдах систем MS ACCESS, Ж. Эрдэнэчимэг, Б. Энхтуул, УБ хот, 2009 он
5. Магадлал математик статистикийн аргыг инженерийн мэргэжилд хэрэглэх, Д.Нямсүрэн, Улаанбаатар
6. Database Concepts, Kroenke, David M, David J. Auer.
7. Microsoft SQL Server 2008 Data Mining Add-ins програмын тусламжийн файл
8. Г. Дугарчулуун, Г. Очирбат, Мэдээллийн технологийн зарим англи нэр томъёоны тайлбар, Улаанбаатар хот, 2010 он, хуудас 30
9. Data Mining with Microsoft SQL Server 2008 Jamie MacLennan, ZhaoHui Tang, Bogdan Crivat ISBN: 978-0-470-27774-4, USA
10. Data Mining with Microsoft SQL Server 2005 ZhaoHui Tang, Jamie MacLennan, ISBN-13: 978-0-471-46261-3, ISBN-10: 0-471-46261-6, USA
11. Data Mining A TOTURIAL-BASED PRIMER, Richard J. Roiger, Michael W. Geatz, 2003, ISBN 0-201-74128-8, USA
12. <http://www.google.mn/>
13. <http://www.functionx.com/sqlserver/Lesson-E> хичээл
14. <http://www.microsoft.com/sqlserver/2008/en/us/data-mining-addins.aspx>
15. <http://en.wikipedia.org>
16. <http://ieeexplore.ieee.org>
17. http://en.wikipedia.org/wiki/Data_mining

