

# Нийтийн тээврийн автобусны аяллын хугацааг өгөгдлийн шинжилгээний аргаар урьдчилан таамаглах нь

Чинзориг Билгүүн

Электроник Холбооны инженерчлэлийн тэнхим  
Монгол Улсын Их Сургууль – ХШУИС  
Улаанбаатар хот, Монгол улс  
[ulaan\\_rockr@yahoo.com](mailto:ulaan_rockr@yahoo.com)

Ганболд Амарсанаа

Мэдээлэл, Компьютерийн ухааны тэнхим  
Монгол Улсын Их Сургууль – ХШУИС  
Улаанбаатар хот, Монгол улс  
[amarsanaag@num.edu.mn](mailto:amarsanaag@num.edu.mn)

*Хураангуй*— Автобусны буудал хооронд явах хугацааг зөв таамаглах нь ухаалаг нийтийн тээврийн системд тулгардаг нэг асуудал юм. Энэ ажлаар бид автобусны аялалын хугацааны өгөгдлөөс машин сургалтын алгоритмын загвар боловсруулж автобусны буудал дээр очих хугацааг таамаглах зорилготой. Үүний тулд буудал хоорондын нэгж шилжилттэй холбоотой 5 бүлгийн онцлогуудыг тодорхойлж тохирох загварыг гаргаж авсан. Энэ загварыг Улаанбаатар хотын нийтийн тээврийн том оврын автобусны аяллын бүртгэлийн өгөгдөл дээр туршихад дунджаар 90%-ийн нарийвчлалтай зөв таамагласан болно.

*Түлхүүр үг*— их өгөгдөл, машин сургалт, нийтийн тээвэр, өгөгдлийн шинжилгээ

## I. УДИРТГАЛ

Ухаалаг нийтийн тээврийн систем гэдэг нь үйлчилгээг сайжруулах, төлөвлөлтийг хялбарчлах гэх мэт зорилготойгоор мэдээллийн технологийг ашиглаж буй нийтийн тээврийн систем юм. Үүний нэг хэсэг болох буудал хоорондын автобусны аялах хугацааг оновчтой тааварлах нь иргэдэд цагаа зөв төлөвлөх, нийтийн тээврийн компаниуд цагийг илүү оновчтой хуваарилах, хотын замын ачааллыг судлах зэрэг ач холбогдлуудтай.

Энэхүү ажлаар бид нийтийн тээврийн аяллын хугацааг таамаглах загварыг хэрхэн боловсруулах талаар судалсан. Үүний тулд аяллыг зэргэлдээ буудал хоорондын шилжилтүүдэд (Нэгж Шилжилт) хувааж загварт тохирох онцлогуудыг тодорхойлж болон машин сургалтын алгоритмуудыг харьцуулав. Мөн алгоритмын сургалтыг бүлэглэн хувааснаар нарийвчлал хэрхэн сайжирч байгааг харууллаа. Бид судалгаагаа Улаанбаатар хотын Smart Card цахим мөнгө хураах системийн цуглуулсан өгөгдөл дээр хийсэн.

2-р хэсэгт өмнө хийгдэж байсан судалгаануудын дэвшүүлсэн аргууд болон тэдгээрийн үр дүнг харьцуулсан бол 3-р хэсэгт Улаанбаатар хотын өгөгдлийг хэрхэн боловсруулсан болон онцлогуудыг сонгон авсныг дурьдав. Удаагаар 4-р хэсэгт 3-н шат дараалалтайгаар дэвшүүлж буй загвараа боловсруулж 5-р хэсэгт үр дүнг нь ажигласан.

## II. ӨМНӨ ХИЙГДЭЖ БАЙСАН АЖЛУУД

Тээврийн хэрэгсэлд суурилуулсан GPS, цахим төлбөрийн систем болон нийтийн тээврийн компаниудын тусламжтай үүсгэсэн өгөгдөл дээр аяллын хугацаа төдийгүй нийт замын ачааллыг таамаглах загваруудыг [1-7,12-15] ажилуудаар дэвшүүлжээ. Эдгээр ажлууд тус бүр нь “Өгөгдлийн судалгаа” болон “Таамаглах загварыг сонгох” гэсэн хоёр үндсэн хэсгээс бүрдэж байна.

### A. Өгөгдлийн судалгаа

Өмнөх судалгаануудад хамгийн өргөн хэрэглэгдсэн өгөгдөл нь нийтийн болон бусад тээврийн хэрэгслийн ойролцоогоор 30сек-5минутын интервалтайгаар хураасан GPS-ийн өгөгдөл байв. Уг мэдээллийг ашиглан тээврийн хэрэгслийн аяллын хугацаа болон хугацаа, гариг, саруудын хамаарлыг [1,6,7] ажилууд тодорхойлсон бөгөөд ажлын цаг, амралтын өдөр зэргээс хамаарсан ерөнхий хандлагууд ажиглагдсан. Жишээ нь, ажлын цаг эхлэх болон дуусахад замын ачаалал огцом өсдөг бөгөөд, хамгийн их ачаалал 18-19н цагийн үед байдаг бол амралтын өдрүүдэд замын ачаалал харьцангуй жигд байдаг [7] гм.

[1,4,5,8] ажлуудад давхар байгаль цаг уур, томоохон үзвэр үйлчилгээ болон автобусанд буй хүмүүсийн тоог давхар авч үзсэн боловч GPS-ийн өгөгдөл шиг тодорхой ерөнхий хандлага ажиглагдаагүй. Энэ нь судалгаа хийсэн хотуудын онцлогтой холбоотой байх магадлалтай бөгөөд эдгээр ажлуудад сонгон авсан онцлогууд болон аяллын хугацааны хамаарлын талаар дэлгэрэнгүй судалгаа байхгүй байсан.

### B. Таамаглах загварууд

Аяллын хугацааг таамаглах загварыг боловсруулахын тулд түүнийг ашиглах нэгж хэсэг болон таамаглах алгоритмыг сонгох хэрэгтэй.

Нэгж хэсэг гэдэг нь сонгон авсан таамаглах алгоритмын үр дүнгийн газар зүйн хүрээг хэлнэ. Өмнө дурьдсан судалгааны ажилуудад нэгж хэсгийг өгөгдөл болон загварт хэр тохиромжтой байгаагаар нь өөр өөрөөр тодорхойлсон бөгөөд хооронд нь харьцуулалт хийгдээгүй. Жишээ нь [1,2] ажилуудад нэгж хэсэг нь

сонгон авсан автобусны чиглэлийн бүтэн болон хагас тойргийн аялах хугацаа байсан бол, [4] ажилд тодорхой автобусны буудлуудын хооронд, [6] ажилд уулзвар дайраагүй замаар аялах хугацаа зэргээр тодорхойлсон. Харин [7] ажилд нэгж хэсгээс гадна долоо хоногийн өдрүүдээр өгөгдлийг хуваасан бөгөөд амралтын өдрүүдэд машин сургалтын аргууд илүү өндөр үр дүнтэй байгааг баталсан.

Харин аялалын хугацааг тодорхойлоход ашиглагдсан алгоритмуудыг бид 4-н бүлэгт хуваасан бөгөөд тэдгээрийн судалгааны үр дүнг доор дурьдав.

#### 1) Регресс модонд түшиглэсэн аргууд

Энгийн регресс мод (RT)-нь бусад аргуудтай харьцуулахад үр дүн багатай боловч тэдгээрийг олноор нь нийлүүлж ашиглавал нарийвчлал сайжирдаг. Иймд регресс модонд түшиглэсэн Random Forest (RF), Extremely Random Forest (ET), Gradient Boosting (GB), Adaptive Boosting (AB) гэх мэт алгоритмуудыг [3,7] ажилуудад авч үзсэн. Жишээ нь, RF алгоритмыг SVR-тэй тодорхой автобусны чиглэлийн бүтэн тойргийн өгөгдөлийн хувьд харьцуулахад дундаж квадрат алдаа нь ойролцоогоор 2 дахин бага байсан [7]. Нэгж хэсгээ буудлуудаар тодорхойлсон [3] ажилд, зөвхөн энэ бүлгийн алгоритмуудыг хооронд нь харьцуулжээ. Үр дүнд нь GB дунджаар 20.46%-ийн абсолют алдаатайгаар хамгийн өндөр үзүүлэлтэй байсан ба аялалын урт ихсэх тусам уг алдаа нь 60%-аас экспоненциал хуулиар 11% хүртэл буурсан. Энэ нь бусад ажилуудад үзүүлсэн үр дүнгүүдтэй ойролцоо боловч бусад бүлгийн аргуудтай хийсэн харьцуулалт бага байна.

#### 2) Бусад өгөгдөлд түшиглэсэн аргууд

Машин сургалтын аргууд дундаас Artificial Neural Network (ANN), Support Vector Regression (SVR), k-Nearest Neighbor (KNN) зэрэг алгоритмуудыг хамгийн өргөн ашигласан. Учир нь, эдгээр аргууд нь шугаман бус хамаарлыг зөвхөн өгөгдөл дээр тулгуурлан харьцангуй сайн тодорхойлдог.

KNN алгоритм нь сургалтын өгөгдөл дээр шууд таамаглал хийдэг учир бусад аргууд шиг сургах шаардлагагүй боловч шуугианд мэдрэг. Иймээс өгөгдлийг шуугианаас салгах спектрийн анализ, алдаа үүсгэж буй өгөгдлийг арилгах, өгөгдлүүдийг зайгаар нь жинлэх зэрэг хэд хэдэн сайжруулах оролдлогуудыг хийсэн [12]. Мөн k-ийн утга 10 байхад алгоритм хамгийн өндөр үр дүн үзүүлсэн [11].

ANN-нь тархины үйл ажиллагаа дээр суурилсан алгоритм бөгөөд шуугианд тогтвортой ажилладаг учир [2,4,6,8]- ажилуудад үндсэн загвараар ашиглагдсан байна. Зөвхөн GPS-ийн өгөгдөл дээр судалгаа хийсэн [6] ажилд ANN-нь Kalman filter болон Historic Average- аргуудаас алдааны стандарт хазайлт болон хувиараа бага байсан. Мөн ANN-г загварчилах ерөнхий хууль байдаггүй бөгөөд хар хайрцаг болгон авч үздэг тул тусгайлан эмпирик судалгаа хийх шаардлагатай. Иймд [2] ажилд хэд хэдэн ANN-ийн загваруудыг хооронд нь харьцуулсан бөгөөд 5-н нуугдмал неоронтой тангенс сигмоид функцтэй загвар 6-10% ийн алдааны нарийвчлалтайгаар хамгийн өндөр үр дүн үзүүлсэн. Мөн [13] ажил ANN-ээр нийлмэл загвар үүсгэж илүү өндөр үр дүнд хүрсэн боловч тооцоолол хийхэд хүндэрсэн.

SVR нь Support Vector Machine-ий дискрет бус үр дүнг тааварлахад зориулсан хувилбар бөгөөд ANN-оос хэд хэдэн давуу талуудтай. Энэхүү алгоритмыг [7] ажилуудад Random Forest болон Historical Average аргуудтай харьцуулсан бөгөөд харьцангуй муу үр дүн үзүүлсэн. Харин [1] ажилд стандарт хазайлт ихтэй өгөгдөлд SVR-нь хамгийн тохиромжтой гэсэн дүгнэлтэд хүрсэн боловч зөвхөн 3-р бүлгийн алгоритмтай харьцуулсан. Өмнө дурьдсан энэ бүлгийн 3-н аргууд дээр нарийн судалгаа хийхэд сургалтын параметруудийг зөв тааруулсан SVR-нь 7 хоног болон түүнээс дээш өгөгдлийн хэмжээнд хамгийн өндөр буюу 3.25км аялалын хугацааг 73-н секундын алдаатайгаар тааварлажээ [11].

#### 3) Дунджид түшиглэсэн аргууд

Өмнөх ажлуудад замын ачаалал нь хугацаанаас ихээхэн хамаардаг бөгөөд ижил хэв маягууд цаг, өдөр, 7 хоног, сарын үетэй ажиглагдсан. Иймээс тухайн агшин дахь хугацааны мэдээллийн дагуу цуглуулсан өгөгдлийн бүртгэлээс аяллын хугацааг тааварлах нь энгийн хэрнээ оновчтой шийдэл болох боломжтой. Энэхүү бүлэгт дээрхи үндэслэлийн дагуу өмнөх өгөгдлийн дунджийг ашигладаг аргууд багтах бөгөөд өгөгдлийн бүртгэлээс гарган авч буй мэдээллээр нь хурдны болон аялалын хугацааны хэмээн ангилж болно. Өгөгдлийн бүртгэлийн дунджаас аялалын хугацааг таамаглах оролдлогууд ихэвчлэн өөр алгоритмуудтай харьцуулахын тулд [8] хийгдсэн бөгөөд зарим тохиолдолд Linear Regression (LR)-ээс илүү үр дүн үзүүлсэн. Мөн real-time GPS-ийн өгөгдөл ашиглан аялалын хугацааг харьцангуй сайн тааварлах боломжтой [9,14]. Харин [5] ажилд дээрхи хурд болон аялалын хугацааны аргыг хослуулан нийт өгөгдлийг variance-ийн дагуу бүлэглэж KNN-ээс илүү үр дүнд хүржээ. Гэвч энэ бүлгийн аргуудыг гадны нөлөө бага, real-time өгөгдөлтэй нөхцөлд л ашиглах боломжтой бөгөөд ихэнхи судалгаанд хамт харьцуулагдсан аргуудаас нарийвчлалаараа бага байсан.

#### 4) Kalman шүүлтүүр

Kalman шүүлтүүр нь шуугиантай өгөгдөлд оновчтой загварыг хурдан олоход хэрэглэгдэх бөгөөд өмнөх болон одоогийн төлвөөс ирээдүйн утгуудыг таамагладаг. Энэхүү аргыг real-time GPS өгөгдөл дээр ашиглан автобусны аялалын хугацааг оновчтой тодорхойлох боломжтой [8]. Зарим тохиолдолд LR болон 3-р бүлгийн алгоритмуудаас илүү үзүүлэлттэй байсан. Мөн Kalman шүүлтүүрийг ашиглан тээврийн хэрэгслийн аялалыг тааварлах оролдлогууд хийгдсэн боловч ихэвчлэн real-time загвар хийхэд ашиглажээ.

Хэд хэдэн аргуудыг нийлүүлж гибрид загвар үүсгэх оролдлогууд давхар хийгдсэн [10,13,15]. Жишээ нь, SVR-ийг сургаад тестийн үр дүнг ашиглан нийт үзүүлэлтэд сөргөөр нөлөөлж буй өгөгдлийг шүүх, Kalman шүүлтүүрийн шинж чанарыг ашиглан сургалтын өгөгдлийг шуугианыг шүүх гэх мэт [15].

Дээрхи судалгааны ажилуудаас үзэхэд хүнд тооцоолол, өгөгдлийн олдоц болон шуугианы улмаас судалгаа хийсэн өгөгдлийг ихэвчлэн зөвхөн нэг чиглэл, буудал эсвэл өдрөөр хязгаарласан. Мөн загвар тодорхойлоход ашигласан онцлогууд, болон тэдгээрийн ач холбогдлын талаар тусгайлан хийсэн

судалгаа дутагдаж байна. Эцэст нь ихэвчлэн ижил бүлгийн алгоритмуудыг хооронд нь харьцуулсан бөгөөд нэмэлт судалгаа шаардлагатай.

### III. ӨГӨГДӨЛ БОЛОВСРУУЛАЛТ

Бид энэхүү судалгаагаа Улаанбаатар хотын OpenData цахим порталаас нийтийн тээврийн системийн талаархи болон түүнд суурилагдсан Smart Card-системийн 2016 оны 2-дугаар сард цуглуулсан өгөгдөл дээр хийсэн (Хүснэгт1).

Хүснэгт 1. АШИГЛАСАН ӨГӨГДЛҮҮД

Өгөгдлийн нэр	Товч тайлбар
1. Автобусны буудалд ирсэн хугацааны бүртгэл	Автобус тус бүрийн буудал дээр ирсэн цаг болон уг буудлаас автобусанд суусан хүмүүсийн тоо
2. Автобусны чиглэл	Автобусны 86 чиглэлийн ID болон нэр
3. Автобусны чиглэлийн газарзүйн зураг	Автобусны 86 чиглэлийн KML өргөтгөлтэй газарзүйн зураг
4. Автобусны буудал	Автобусны 723 буудлуудын ID, координат, нэр
5. Автобусны чиглэлийн буудлуудын дугаарууд	Автобусны чиглэл тус бүрийн дайрч өнгөрөх буудлуудын ID
6. Хур тунадасны мэдээлэл	Улаанбаатар хотын 2016 оны 2 сарын хур тунадасны мэдээлэл

Бидний ашиглаж буй өгөгдлийн хүрээ 5-6-н буудал эсвэл 1 чиглэлийг авч үзсэн [1-7,9-10,12-14] ажилуудаас хэт их байгаа бөгөөд ижил таамаглах арга ашиглах нь тохиромжгүй гэж үзэж байна. Иймд [3] ажлаас санаа авч нийт өгөгдлийг нэгж шилжилт (НШ) буюу автобусны чиглэлийн дагуу зэргэлдээ буудлуудын хоорон дахь шилжилтээр хуваасан бөгөөд нийт аялалын хугацааг эдгээр нэгж шилжилтүүдийн нийлбэр хэмээн тооцов. Ингэснээр, дурын 2 буудлуудын хооронд нийтийн тээврийн аялах хугацааг хялбар бөгөөд илүү нарийвчлалтайгаар тодорхойлох боломжтой.

“Автобусны чиглэл болон буудлуудын дугаарууд” өгөгдлөөс үзэхэд 723 буудлуудын хооронд боломжит 876 нэгж шилжилт байна. Эдгээрийн эхлэх болон төгсөх буудлуудын мэдээлэл дээр тулгуурлан “Автобусны чиглэлийн газарзүйн зураг”-аас бид нэгж шилжилтэд харгалзах замын урт болон траекторын муруйлтын индексийг (өнцгийн өөрчилөлтүүдийн абсолют нийлбэр) олсон. Учир нь аялалын хугацаа нь замын урт болон хэлбэрээс шууд хамаарах магадлалтай. Автобусны буудалд ирсэн хугацааны бүртгэлээс 2016 оны 2 сарын байдлаар нийт 3,391,956 ширхэг нэгж шилжилтүүд бүртгэгдсэн бөгөөд тэдгээрийн аялалын хугацаа болон цагийн мэдээллийг цуглуулсан.

Улаанбаатар хотын замын хөдөлгөөнд хур тунадас ихээхэн нөлөө үзүүлдэг байх магадлалтай тул wunderground.com вебсайтаас 2016 оны 2 сарын хур тунадасны мэдээллийг авсан.

Бидэнд байгаа өгөгдлүүд дээр тулгуурлан нийт 4-н бүлгийн 10-н төрлийн үүсмэл болон үүсмэл биш онцлогуудыг гарган.

Хүснэгт 2. СОНГОН АВСАН ОНЦЛОГУУД

Бүлэг	Онцлогийн нэр	Тайлбар
	Байршил	Буудлуудын координат

НШ-ийн эхлэх төгсөх Автобусны буудал	Харгалзах дундаж хугацаа	Тухайн буудалтай холбоотой НШ-үүдийн аялалын дундаж хугацаа
Хугацаа	Өдөр	НШ бүртгэсэн өдөр
	Гариг	НШ бүртгэсэн гариг
	Цаг	НШ бүртгэсэн цаг
Нэгж шилжилт	Бүртгэсэн хугацаа	НШ бүртгэсэн хугацаа
	Урт	Урт
Нэгж шилжилт	Замын муруйлтын индекс	Тректурын өнцгийн өөрчилөлтүүдийн абсолют нийлбэр
	Харгалзах дундаж хугацаа	Тухайн холбоотой бүртгэлүүдийн дундаж хугацаа
Цаг агаар	Хур тунадас	Тухайн өдөр эсвэл өмнөх өдөр хур тунадас орсон эсэх

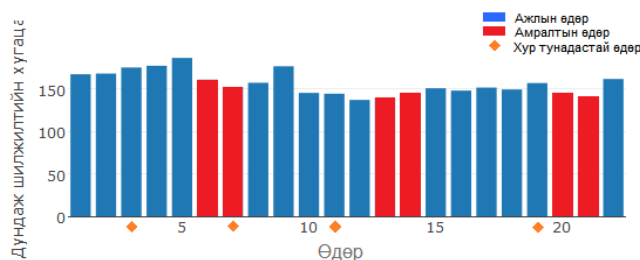
### IV. ЗАГВАР БОЛОВСРУУЛАЛТ

Тохиромжтой загварыг боловсруулахын тулд бид түүнд өгөх хувьсагчууд буюу онцлогуудыг тодорхойлж, ашиглах тохиромжтой алгоритм болон түүний хамрах хүрээг тодорхойлох хэрэгтэй. Энэхүү бүлэг нь гурван хэсгээс тогтоно. Эхний хэсгээр бид хэрхэн сонгон авсан онцлогуудад шүүлт хийснээ, хоёрдугаар хэсгээр харьцуулалтад ашигласан алгоритмуудын талаархи мэдээллийг, сүүлийн хэсэгт загварын тоог ихэсгэхийн ач холбогдлын талаар тус тус дурьдана.

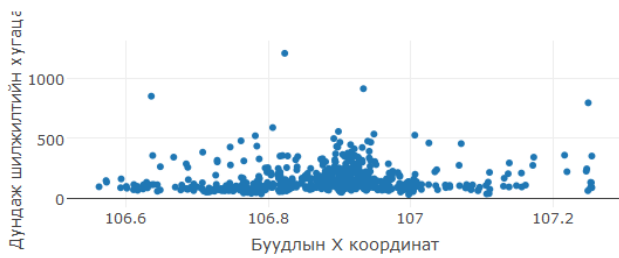
#### A. Онцлогуудын судалгаа

Бидний сонгон авсан онцлогуудын ач холбогдлыг тодорхойлохын тулд тус бүрийн аялалын хугацаанд үзүүлэх нөлөөлийг судлах хэрэгтэй. Үүний тулд бид тухайн онцлогийн утга болон түүнд харгалзах нэгж шилжилтүүдийн дундаж аялалын хугацааны графикийг байгуулсан (Жишээ: Зураг 1,2). Эдгээр графикуудаас бид дараахь дүгнэлтүүдэд хүрсэн.

- НШ-ийн аялалын хугацаа болон бүртгэсэн хугацааны хамаарал нь [7] ажилын үр дүнтэй төстэй байна.
- Хотын төвөөр аялалын хугацаа эрс нэмэгддэг
- Нэгж шилжилтийн урт, муруйлтын индекс болон аялалын хугацааны хооронд мэдэгдэхүйц шугаман хамаарал байхгүй
- Хур тунадас орсон өдрүүдэд дундаж аялалын хугацаа өмнөх өдрөөс 5-8-н секундээр нэмэгдсэн.



Зураг 1. Өдөр болон дундаж хугацааны хамаарал



Зураг. 2. Буудлуудын байршил болон дундаж хугацааны хамаарал

Сонгон авсан онцлогуудын ач холбогдлыг тодорхойлохын тулд RF алгоритмийг онцлог тус бүрээр дангаар нь сургаж алдааны хувиар нь эрэмбэлсэн. RF алгоритм нь өгсөн онцлогыг ашиглан өгөгдлийг нийт хазайлт нь хамгийн бага байлгах бүлгүүдэд хуваадаг тул аль онцлог нь нарийвчлалд илүү үр нөлөөтэй байгааг тодорхойлоход тохиромжтой.

Нийт ашигласан онцлогууд дунд харгалзах дундаж хугацаа, нэгж шилжилтийн урт зэрэг үүсмэл онцлогууд нь ижил бүлгийн онцлогуудтай давхардсан мэдээлэл агуулж байх боломжтой. Иймд аялалын хугацааг таамаглахад хамгийн тохиромжтой онцлогуудын комбинацыг тодорхойлохын тулд онцлог тус бүрийн ач холбогдлоос гадна хоорондын хамаарлыг давхар тооцох ёстой. Иймд бид корреляцийн матрикс ашиглан ижил төстэй онцлогуудыг олсон.

Эцэст нь дээрхи цуглуулсан мэдээллээ ашиглан онцлогуудыг нэг нэгээр хасж алдааны хувийн өөрчлөлтийг ажиглан хамгийн тохиромжтой онцлогуудыг сонгон авна.

**В. Таамаглах аргыг сонгох**

Энэхүү судалгаагаар бид регресс модонд түшиглэсэн аргуудаас Random Forest болон Gradient Boosting, өгөгдөлд түшиглэсэн аргуудаас Support Vector Regression, дунджид түшиглэсэн аргуудаас Look Up Table зэргийг сонгож хооронд нь харьцуулсан. Ашигласан аргуудын талаархи товч мэдээлэл болон сургалтын параметруудийг хэрхэн тааруулсан талаар доор дурьдав.

**1) Random Forest**

Энэхүү алгоритм нь өгөгдлийг санамсаргүй байдлаар хэд хэдэн хэсэгт хуваагаад нэгж бүрт тохирох регресс мод үүсгээд тэдгээрийн үр дүнгийн дунжийг ашигладаг нийлмэл арга юм. Үүний давуу тал нь их өгөгдөлд ашиглахад тохиромжтой, нарийвчлал сайн, алдагдсан өгөгдлөөс хамаарахгүй бөгөөд онцлогуудын аль нь илүү үр дүнд нөлөөтэй болохыг мэдэх боломжтой зэрэг болно.

Тохируулах параметрууд нь:

1. N: регресс модны тоо
2. ns: модны 1 зангилаа доторхи хамгийн бага өгөгдлийн тоо

**2) Gradient Boosting**

Random Forest-ийн адил регресс моднуудаас бүрдсэн нийлмэл алгоритм боловч, дараа дараагийн моднуудыг нь өмнөх моднуудын алдаан дээр сургадаг. Үр дүнгийн хувьд Random Forest-оос ихэвчлэн илүү боловч амархан overfitting үүсэх хандлагатай байдаг.

Тохируулах параметрууд нь:

1. N: регресс модны тоо
2. ns: зангилаа доторхи хамгийн бага өгөгдлийн тоо
3. shrinkage: сургалтын хурд

**3) Support Vector Regression**

Ашиглаж буй санаа нь Support Vector Machine-тай адил бөгөөд ашиглаж буй хязгаарлалтаараа өөр юм. SVM нь сургалтын параметруудаас ихээхэн хамаардаг бөгөөд over-fitting үүсэх хандлагатай. Гэхдээ тохиромжтой кернел функц болон алдааны жинг олж чадвал харьцангуй өндөр үр дүн үзүүлдэг. Мөн SVM-нь конвекс оновчлолоор тодорхойлогддог тул local minimum гэж байхгүй. Кернел функцийг хувьд [4] ажилын дагуу Radial Basis Function ашигласан.

Тохируулах параметрууд нь:

1. C: алдааны жин
2. gamma: кернел функцд ашиглагдах параметр

**4) Look up table**

Энэ нь хугацаа болон нэгж шилжилтийн дагуу өгөгдлийг стандарт хазайлт нь тодорхой утгаас бага байхаар бүлэглээд, тэдгээрийн аялалын хугацааны дунджаар үүсгэсэн хүснэгт юм. Уг аргыг хэрэглэхэд хялбар боловч overfitting үүсэх хандлагатай.

Тохируулах параметрууд нь:

1. V: Нийт стандарт хазайлтын болон нэг бүлэгт зөвшөөрөгдөх максимум стандарт хазайлтын харьцаа.
2. ms: Нэг бүлэгт агуулагдах хамгийн бага бүртгэлийн тоо

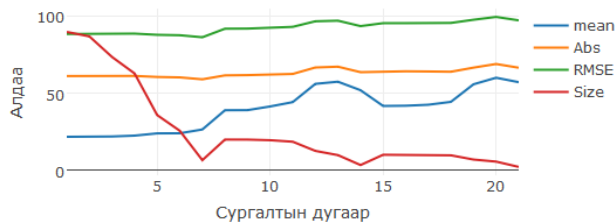
**5) Сургалтын параметруудыг тохируулах нь**

Бидний ашиглаж буй эдгээр аргууд нь бүгд өгөгдлөөс хамаардаг. Нэгэнт эдгээрийн сургалтын параметруудыг тухайн асуудлын хувьд сонгох ерөнхий дүрэм байдаггүй тул эдгээрийг эмпирик аргаар олдог [2,8]. Мөн цаашлаад тухайн сургалтын өгөгдөл бүрт хамгийн тохиромжтой параметрууд нь өөр өөр утгатай байдаг. Бид тухайн алгоритм бүрд зөвхөн нэг параметруудыг өөрчилж нарийвчлал хамгийн өндөр байгаа утгыг тэмдэглэх замаар хамгийн тохиромжтой утгуудыг тодорхойлсон. Жишээ болгож LUT алгоритмийн хувьд сургалтын параметруудыг хэрхэн тодорхойлсныг авч үзэв.

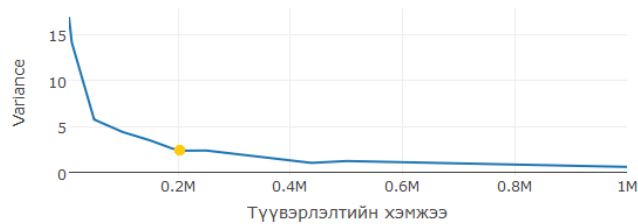
V параметрийн хувьд 7, ms параметрийн хувьд 3-н ширхэг боломжит утгуудыг судалсан бөгөөд нийт 21-н сургалт явуулсан. Сургалт тус бүрийн дундаж алдаа, абсолют алдаа, дундаж квадратлаг алдаа болон хүснэгтийн хэмжээг мегабайтаар зураг 3-т харуулсан бөгөөд ms=1 болон V=10000 байх нь хамгийн оновчтой.

ХүСНЭГТ 3. АШИГЛАСАН ӨГӨГДЛИҮҮД

Сургалтын Параметрууд	Хайлтын утгууд
V	5,10,50,100,500,1000,10000
Ms	1,5,10



Зураг 3. RF алгоритмын модны тоо болон алдааны хувийн хамаарал



Зураг 4. Стандарт хазайлт болон түүвэрлэлтийн хэмжээ

**С. НШ тус бүрт загвар боловсруулах**

Бидний сургалтын өгөгдлийн хамрах хүрээ [] ажилуудынхаас харьцангуй их бөгөөд стандарт хазайлт нь их байгаа тул таамаглах аргад муугаар нөлөөлж болзошгүй. Ялангуяа хотын зах болон чиглэлийн эцсийн буудлуудад аялалын хугацааны стандарт хазайлт их байсан. Иймд нэгж шилжилт тус бүрт загвар боловсруулснаар нийт үр дүн сайжрих боломжтой.

**V. ТУРШИЛТ**

Бид туршилтаа 4-р бүлгийн дагуу онцлогуудыг судлах, таамаглах аргыг сонгох, НШ тус бүрт загвар боловсруулах гэж 3 хуваасан. Энэхүү хэсэгт бид туршилтын өгөгдлийг хэрхэн боловсруулсан болон үр дүнг дурьдана.

**A. Туршилтын өгөгдөл**

Харьцуулалтад сонгон авсан алгоритмууд нь сургалтын болон шалгах гэсэн 2 төрлийн өгөгдлөөс бүрддэг. Эдгээрийг бүлгүүдийг бүрдүүлэхэд ихэвчлэн k-fold cross validation буюу өгөгдлийг k бүлэгт хуваагаад k-1 хэсгийг сургалтад ашигладаг. Энэхүү процессийг k дахин давтахад бүхий л өгөгдөл сургалтад болон тест-д ашиглагдах боломжтой. Гэвч бидний ашиглаж буй өгөгдөл нь хэт их бөгөөд SVR зэрэг алгоритмуудыг сургахад хугацаа өгөгдлийн тооны 2.3-н зэрэгтээр хамаарадаг тул 2 бүлгийг дараахь аргаар хувааж ашигласан.

**1) Сургалтын өгөгдөл**

Бидэнд нийт 3-н сая гаруй боломжит сургалтын өгөгдөл байгаа бөгөөд SVR зэрэг алгоритмыг сургах хугацаа нь өгөгдлийн тооны куб зэргээс хамаарадаг тул нийт сургалтын өгөгдлийг илэрхийлж чадахуйц хамгийн бага түүвэрлэлтийн тоог тогтоох хэрэгтэй. Ингэснээр туршиж буй алгоритмуудыг нарийвчлал буурах боловч сургах хугацаа нь хурдсах юм. Үүний тулд бид өгөгдлөөс 10,000 аас 1 сая хүртэл 15-н өөр төрлийн хэмжээтэй нийт 150н бүлэг үүсгээд бүлэг тус бүрийг энгийн регресс мод ашиглан сургаад алдааны хувийн хамаарлыг судалсан. Ингэхдээ, түүвэрлэлтийн утга ихсэх тусам ижил хэмжээтэй бүлэг дээр сургасан регресс модны алдаануудын стандарт хазайлт нь буурах бөгөөд энэ нь нийт өгөгдлийг хэр сайн илэрхийлж байгааг илтгэнэ гэж үзэв. Зураг4-т түүвэрлэлтийн стандарт хазайлт болон хэмжээний хамаарлыг харуулсан бөгөөд хамгийн тохиромжтой сургалтын өгөгдлийн түүвэрлэлтийн утга нь 200,000 байна.

**2) Шалгах өгөгдөл**

Улаанбаатар хотын чиглэлүүд нь дунджаар 1 талдаа 24-н буудалтай буюу 23-н нэгж шилжилттэй. Иймд бид дэвшүүлж буй аргаа шалгахын тулд 1 ээс 23 н ширхэг үргэлжилсэн нэгж шилжилтээс тогсон 50000 аялалын бүртгэлийг автобусны буудалд ирсэн хугацааны бүртгэлээс гарган авсан. Дараа нь аялал тус бүрийн нэгж шилжилтүүдийг таамаглаад тэдгээрийн нийлбэр болон бодит аялалын хугацаа хоёрыг томъёо 1 ашиглан дундаж алдааны хувийг (ДАХ) тодорхойлсон.

$$ДАХ = \frac{\sum_n \left| \frac{\sum_m P_{nm} - A_n}{A_n} \right|}{N} \quad (1)$$

Үүнд P-нь таамагласан нэгж шилжилтийн, A-нь бодит аялалын хугацааны утгуудын бол M-нь тухайн аялалд хамаарагдах НШ-ийн, N-нь аяллын тоо юм.

**B. Үр дүн**

Бид 3-н төрлийн туршилт явуулсан тул үр дүнг ч бас 3 хуваасан. Эхний хэсэгт онцлогуудын ач холбогдол, ижил төстэй байдал болон тэдгээрийн үр дүнтэй хамаарах хамаарлыг аялалын урт нь 1 НШ байх өгөгдөл дээр тулгуурлан судалсан. Харин, хоёрдугаар хэсэгт 4.В.5-р бүлэгт тодорхойлсон сургалтын параметруудыг ашиглан сургалт болон шалгах өгөгдөл ашиглан алгоритмуудад үнэлгээ хийж харьцуулав. Эцэст нь НШ тус бүрт загвар боловсруулахдаа сургалтын өгөгдөлд нийт 3 сая бүртгэл болон шалгах өгөгдлийг ашигласан. Учир нь ганц НШ-д харгалзах өгөгдлийн хэсэг нь харьцангуй бага учир хотын захын нэгж шилжилтүүдийн хувьд үнэлгээ хийхэд хангалттай өгөгдөл байхгүй байна. Мөн нэгж шилжилтийн хувьд байршлийн мэдээлэл тогтмол учир зөвхөн хугацааны онцлогуудыг ашигласан.

**1) Онцлогуудын судалгаа**

Онцлог хоорондын коррелиацийн матрицийн дундаж утга нь 0.16 буюу харьцангуй буюу онцлогууд хоорондоо төсөө багатай байна. Бүлэг хоорондын онцлогуудын коррелиацууд нь маш бага буюу -0.007 оос +0.008 байна. Хүснэгт 4-д 0.5 аас их коррелиацийн утгуудтай хос онцлогуудыг харуулав.

Хүснэгт 4. Өндөр коррелиацтай онцлогууд

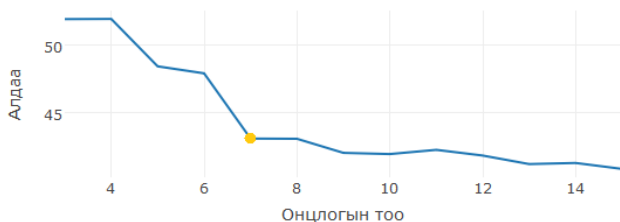
Коррелиацийн утга	Онцлог №1	Онцлог №2
0.821	НШ-ийн эхлэх буудал	НШ-ийн төгсөх буудал
0.999	Бүртгэсэн хугацаа	Өдөр
0.742	НШ-ийн урт	НШ-ийн муруйлтийн индекс

Онцлогууд тус бүрээр дангаар нь сургасан DT-ийн утгуудыг ашиглан онцлогуудыг эрэмбэлсэн бөгөөд энэ нь тус бүрийн үр дүнд нөлөөлөх ач холбогдлыг илэрхийлнэ. Хүснэгт 5-д тус бүрийн ДАХ болон онцлогуудыг харуулав

Хүснэгт 5. Онцлог болон ДАХ-ийн хамаарал

ДАХ	Онцлог	ДАХ	Онцлог
33%	НШ-ийн дугаар	35.7%	НШ-ийн координат
33.2%	НШ-ийн урт	49.4%	Бүртгэсэн хугацаа
34.6%	НШ-ийн эхлэх буудал	53.4%	Өдөр
35.4%	НШ-ийн төгсөх буудал	53.6%	Цаг
35.5%	НШ-ийн координат	53.8%	Гариг
35.6%	НШ-ийн муруйлт	53.8%	Цаг агаар

Ижил 0.5-аас их коррелиацтай онцлогуудаас хамгийн бага ДАХ-тай утгуудыг, дараагаар нь онцлогуудыг эрэмбийн дагуу нэг нэгээр нь хасаад DT-г дахин дахин сургасан үр дүнг зураг5-т харуулав.

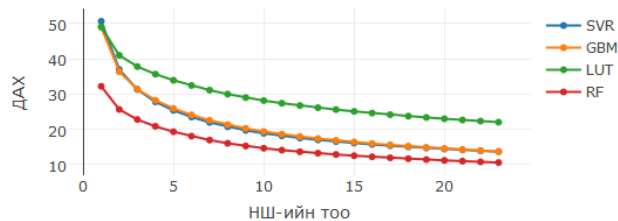


Зураг 5. DT алдаа болон онцлогуудын тооны хамаарал

Харин хугацаатай холбоотой онцлогуудын хувьд бүртгэсэн хугацаа нь бусад 3-н онцлогуудынхаа мэдээллийг давхар агуулж байгаа бөгөөд Өдөр, Цаг, Гаригуудыг хамт ашигласнаар 3.3%-аар илүү үр дүнд хүрсэн. Энэ мэтээр эцэст нь бид 2% ийн алдааны өсөлттэйгөөр 8-н онцлогыг хассан бөгөөд үүнээс олныг хасах нь алдааны хувийг хэт ихэсгэж байна.

2) Таамаглах аргыг сонгох

Өмнөх туршилтын хэсгээр тохиромжтой онцлогуудыг тодорхойлсон тул уг үр дүнг ашиглан сонгон авсан алгоритмуудаа харьцуулсан. Үүний тулд сургалтын өгөгдлүүдийг ашиглан алгоритм тус бүрийг сургалтын хамгийн тохиромжтой параметруудыг ашиглан нэг удаа сургасан. Харин үнэлгээ хийхдээ шалгах өгөгдлийн бүтэн аялал тус бүрийн утгыг түүнд харгалзах НШ-үүдийн аялалын хугацааг таамаглаж нэмэх замаар гарган аваад бодит үр дүнтэй харьцуулсан алдааны хувийг ашигласан. Зураг6-т алгоритм тус бүрийн ДАХ болон аялалын НШ-ийн тооны хамаарлыг харуулав



Зураг 6. Аяллын урт болон алгоритмуудын алдааны хувийн хамаарал

Мөн эдгээр алдааны хувиудийн дундаж утгуудыг хүснэгт6-д харуулсан бөгөөд RF алгоритм нь харьцангуй өндөр үзүүлэлтэй байна.

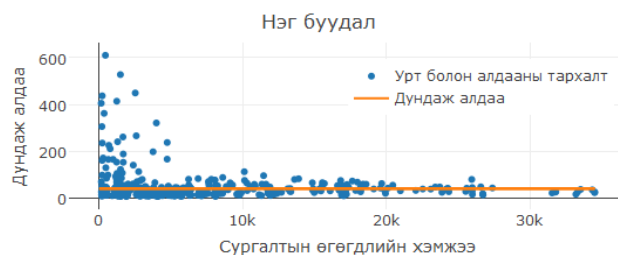
Хүснэгт 6. Алгоритмуудын ДАХ-ийн дундаж

НШ-ийн тоо	1	5	10	15	20
SVR	50.6	25.38	18.89	16.12	14.46
GBM	48.94	25.93	19.4	16.44	14.58
LUT	49.02	33.9	28.16	25.09	23.01
<b>RF</b>	<b>32.2</b>	<b>19.34</b>	<b>14.66</b>	<b>12.55</b>	<b>11.23</b>

3) НШ тус бүрт загвар боловсруулах

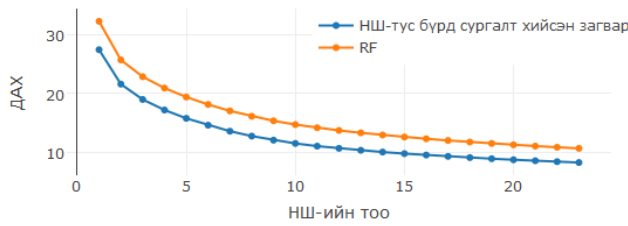
Бид энэ хэсэгт, өмнөх туршилтийн хэсгийн үр дүн дээр тулгуурлан зөвхөн RF алгоритмийг ашигласан. Учир нь уг алгоритм нь харьцангуй өндөр нарийвчлалтай байснаас гадна сургах хугацаа нь SVR-ээс харьцангуй бага юм.

Нийт сургалтын өгөгдлийг НШ-үүдээр нь хуваахад зарим бүлэг 100-аас доош бүртгэлтэй буюу сургалт хийхэд хангалттай бус байсан. Мөн загвар тус бүрийн үр дүн сургалтын өгөгдлийн хэмжээнээс их хамаарч байсан бөгөөд зураг7-т энэхүү хамаарлыг харуулав.



Зураг 7. Сургалтын өгөгдлийн хэмжээ болон загварын дундаж алдаануудын хамаарал

Дараагаар нь бид өмнөх туршилтын хэсэгт хийсний адилаар ДАХ болон аялалын НШ-ийн тооны хамаарлыг энэхүү арга (RF+)-ийн хувьд судалж энгийн RF загвартай харьцуулсан (Зураг8, Хүснэгт 7).



Зураг. 8. Аялалын урт болон ДАХ-ийн хамаарал  
Хүснэгт7 Алгоритмуудын ДАХ-ийн дундаж

НШ-ийн тоо	1	5	10	15	20
RF+	27.38	15.72	11.46	9.73	8.69
RF	32.2	19.34	14.66	12.55	11.23

Дээрхи үр дүнгээс үзэхэд өгөгдлийг хувааснаар хэдий загварын тоо ихсэх боловч RF алгоритмийн үр дүнг дунджаар 28% аар илүү нарийвчлалтай болгох боломжтой юм.

### ДҮГНЭЛТ

Энэхүү ажилаар бид хотын хэмжээнд автобусны дурын буудал хоорондын аялалын хугацааг тодорхойлох загварыг боловсрууллаа. Ингэхдээ өгөгдлийг нэгж шилжилтүүдийн дагуу хувааж боломжит 15-н онцлогуудыг тодорхойлсон. Дараа нь, өгөгдөлд тохирсон аялалын хугацааг таамаглах загварыг боловсруулахын тулд тохиромжтой 7-н онцлогыг илрүүлж, машин сургалтын алгоритмуудыг харьцуулж, загварын тоог ихэсгэсэн. Үүний үр дүнд RF алгоритм нь харьцангуй өндөр үр дүн үзүүлснээс гадна нэгж шилжилт тус бүрт загвар боловсруулснаар дундаж алдааны хувь нь өмнөх ажилуудынхаас № хувиар буурсан

Өгөгдлийг нэгж шилжилт бүрээр нь хуваасан тул бид нэг дор 876-н RF загваруудтай харьцах хэрэгтэй болсон бөгөөд энэ нь алгоритмийн хурдыг саарууж санах ойд эзлэх хэмжээг ихэсгэсэн. Харин [7] ажилд байршил бус хугацаагаар нь өгөгдлийг хувааж нарийвчлалыг сайжруулсан. Эндээс бид хамгийн цөөн тооны боловч оновчтой хуваалт нь өөр онцлог эсвэл онцлогуудын комбинацын хувьд оршин байх боломжтой гэж үзэж байна. Иймд цаашид тэрхүү хуваалтыг судалж, загварын тоог алдааны нарийвчлалыг ихэсгэлгүйгээр багасгах боломжуудыг судлах хэрэгтэй.

### НОМ ЗҮЙ

- [1] Kranthi Kumar Reddy, B. Anil Kumar and Lelitha Vanajakshi, "Bus travel time prediction under high variability conditions" CURRENT SCIENCE, VOL. 111, NO. 700 4, 25 AUGUST 2016
- [2] Johar Amita1, Jain Sukhvir Singh2, Garg Pradeep Kumar3, "PREDICTION OF BUS TRAVEL TIME USING ARTIFICIAL NEURAL NETWORK", International Journal for Traffic and Transport Engineering, 2015, 5(4): 410 – 424
- [3] Avigdor Gala, Avishai Mandelbaum, Francois Schnitzler, Arik Senderovitch, Matthias Weidlich, "Traveling Time Prediction in Scheduled Transportation with Journey Segments", February 22, 2016
- [4] Erlend Dahl, Aleksander Aas Sjøfjell, Simen Skogen, "On Implementations of Bus Travel Time Prediction Utilizing Methods in Artificial Intelligence", June 2014, Norwegian University of Science and Technology
- [5] Hao Chen, Hesham A.Rakha, Catherine C.McGhee, "Dynamic Travel Time Prediction using Pattern Recognition", 3500 Transportation Research Plaza, Blacksburg, VA 24061
- [6] Wei Fan, Zegeye Gurmu, "Dynamic Travel Time Prediction Models for Buses Using Only GPS Data", International Journal of Transportation Science and Technology . vol 4.no. 4· 2015 - pags 353 – 366
- [7] Joˆao M. MOREIRA, Alˆipio JORGE, Jorge Freire de SOUSA, Carlos SOARES, "TRIP TIME PREDICTION IN MASS TRANSIT COMPANIES. A MACHINE LEARNING APPROACH", Advanced OR and AI Methods in Transportation
- [8] Mehmet Altinkaya, Metin Zontul, "Urban Bus Arrival Time Prediction: A Review of Computational Models", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-4, September 2013
- [9] Farhana Siddiqui, Abbasali Springwala, Danish Shaikh, Shumaila Siddiqui, "GPS Based Bus Arrival Time Prediction System", International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 2, February – 2014 IJERT/IJERT ISSN: 2278-0181
- [10] Shquan Zhong, Juanjuan Hu, Shuiping Ke, Xuelian Wang, JingHian Zhao, Baozhen Yao, "A Hybrid Model Based On Support Vector Machine For Bus Travel-Time Prediction", Intelligent Transport Systems (ITS), Original Scientific Paper, July 8, 2015
- [11] Aleksander Sjøfjell, Erlend Dahl and Simen Skogen, "Intelligent Transportation Systems and Artificial Intelligence – a State of the Art Review", DECEMBER 2013
- [12] T. Liu, J. Ma, W. Guan, Y. Song, and H. Niu, "Bus arrival time prediction based on the k-nearest neighbor method," in Computational Sciences and Optimization (CSO), 2012 Fifth International Joint Conference on, 2012, pp. 480–483.
- [13] C. van Hinsbergen, A. Hegyi, J. van Lint, and H. Van Zuylen, "Bayesian neural networks for the prediction of stochastic travel times in urban networks," Intelligent Transport Systems, IET, vol. 5, no. 4, pp. 259–265.
- [14] S.I.J Chien, and C.M. Kuchipudi, "Dynamic Travel Time Prediction with Real-Time and Historic Data.," Journal of Transportation Engineering, 2003, Volume 129, Number 6, pp. 608-61
- [15] Cong Bai, Zhong-Ren Peng, Qing-Chang Lu, Jian Sun, "Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes", Computational Intelligence and Neuroscience, Jul 30 2015: 432389.