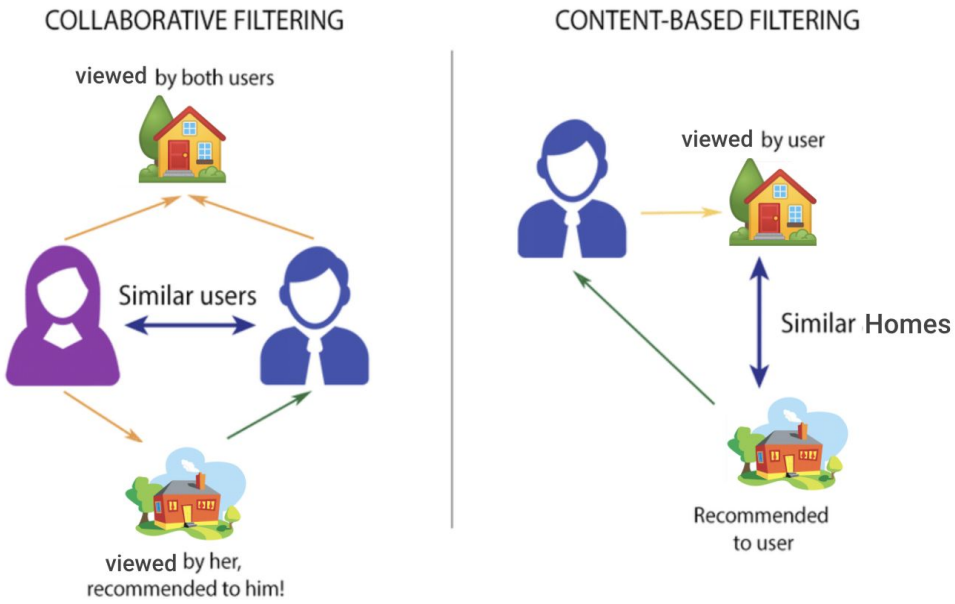
Release Announcement: Machine Learning - Recommended Homes

Recommendations and Data science teams are excited to announce that our newest machine learning model, Recommended homes, was released recently in production.

Recommended homes (RH) is our first fully personalized machine learning (ML) model that has been developed to generate personalized recommendations. The goal is to serve curated lists of unseen houses to users to reduce their decision fatigue and increase lead generation. RH has been optimized for explicit feedback (lead, save, share) and implicit feedback (dwell-time) from user-listing interactions.

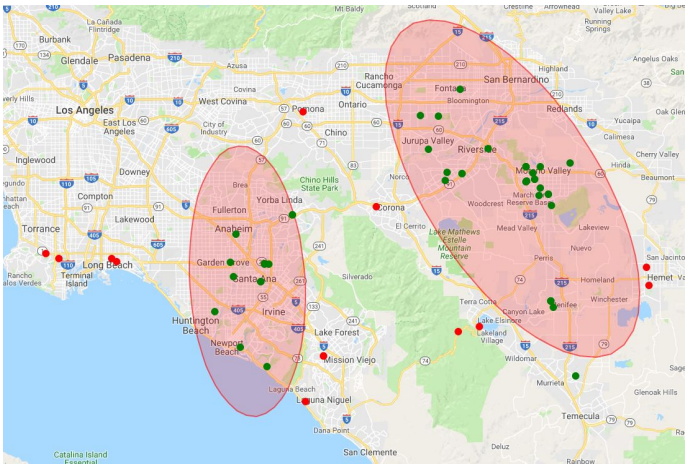
RH has been developed on a hybrid recommendation algorithm using the library LightFM. It is hybrid in the sense that it combines the two fundamental algorithms for recommender systems - collaborative and content-based filtering. This is achieved by using a latent factorization model, which uses listing features to generate better recommendations.



Collaborative Filtering and Content-based filtering

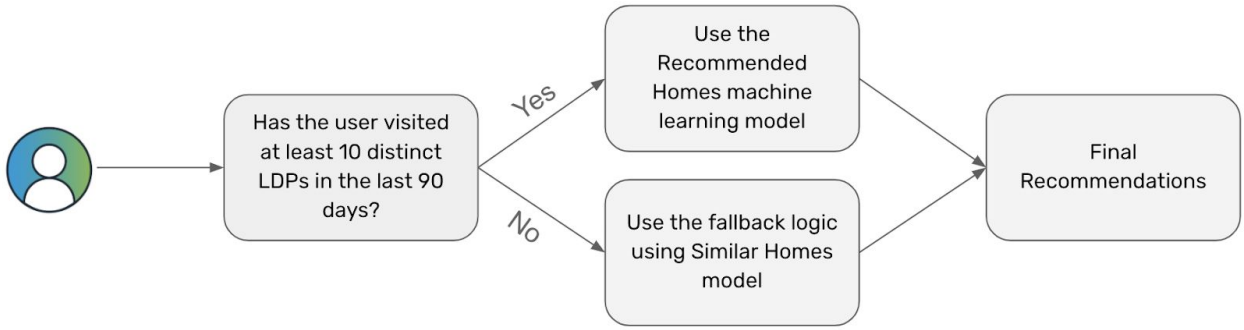
Both explicit and implicit feedback from user-listing interactions are combined to get a hierarchical nonparametric statistical score which is normalized between 0 and 1 for every user. Mathematical operations are performed on listing features to make the model learn from users'

preferences. Also, to constrain the recommendations within a user's preferred geography, we use the density based clustering algorithm DBSCAN on the latitude and longitude of all the houses that user visited.



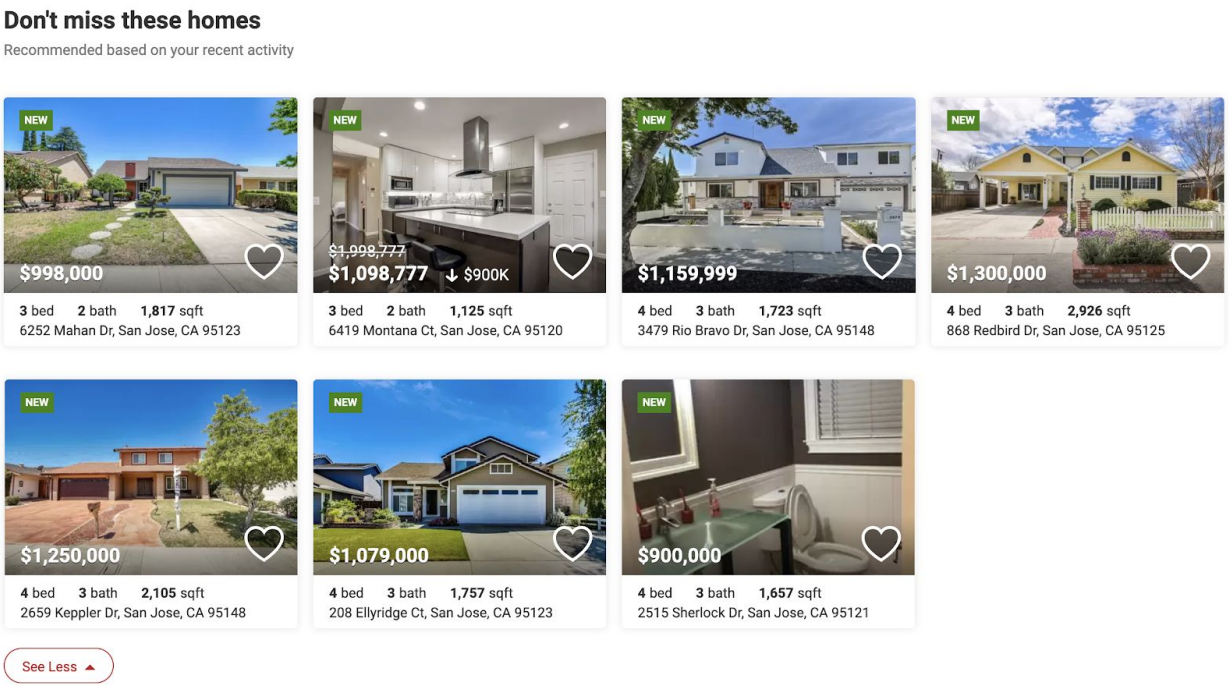
User's preferred geography determined using a clustering algorithm

RH ML model generates recommendations for ~16 million unique users every day in batch, users who have visited at least 10 distinct LDPs in the last 90 days. For other users who have less than 10 distinct LDP views, a fallback logic is implemented to keep the experience consistent amongst all users. The fallback logic generates content-based recommendations using the Similar Homes of the recently viewed listings by the user.



First use case and A/B test results

For the first use case, we A/B tested the RH on the homepage. The RH model is the brain behind the "Don't miss these homes" module on the homepage.



The result of this test was very positive and promising 🥳

Lift (AB test)	Web (desktop)	Mobile Web	Overall
Leads	+2.91%	+4.58%	+3.48%
ULS	+1.20%	+2.93%	+1.70%

*mWeb results are stat-sig at 96% and desktop results are stat-sig at 90%

Next steps

- Candidate generation and model Improvements: We will continue to improve our candidate generation logic and recommendations models, in the future.
- Notifications: We are currently working with the Notifications team to test the recommended homes.

Release Announcement: Machine Learning - Smart Tags

Data Science and Listing Insights teams are excited to announce that our newest machine learning model, Smart Tags, was released recently in production.

Smart Tags is a deep learning-based approach (DistilBERT) that uses natural language processing (NLP) techniques to extract the following tags from listing description:

- Pool: Private Pool, Community Pool, and No Pool
- Basement
- Garage
- Waterfront
- Stories: 1 Story, 2 Story, and 3 or more Stories

These tags are used for populating the results of the corresponding search filters.

Smart Tags is an enhancement to an existing system that was entirely based on regular expressions (regex). This system produces about 106 unique tags that are part of search filters and also searchable via free text search. Out of these 106 tags, we identified 5 tags that are searched the most by our users and we also received the most number of consumer complaints for those tags. The goal was to increase the accuracy of those tags using machine learning; hence, reduce the number of complaints received and improve the relevance of search results.

Regex based solutions use pattern matching to identify the presence/absence of words or phrases. They are not capable of understanding words and phrases taking into consideration syntax, grammar and ultimately context. On the other hand, AI solutions using NLP and Machine Learning can conduct an intelligent analysis of the unstructured text, and in practical terms “read” a document similar to a human - only hundreds of times faster. For Smart Tags, we fine-tuned DistilBERT based language models for the specific use-cases by leveraging listing descriptions and the default tags from MLS’s.

To learn more about Smart Tags click here.

Results

Overall offline testing showed that there was an increase of **5-40%** in the accuracy of the tags from baseline regex. We saw a significant increase in accuracy particularly in the pool and waterfront tags.

The comparison in **precision** for the different tags is shared below. Precision is defined as the percentage of the predictions correctly made by the model, i.e. if the model predicted that 100 listings had a private pool but only 95 listings among them had a pool, then the precision would be 95%.

Precision				
Tag	**Category**	**Baseline Regex**	**SmartTags**	**Change**
Pool	**Private Pool**	61.9%	95.1%	33.2%
	Community Pool	89.0%	92.0%	3.0%
Garage	**Garage: Yes**	97.0%	98.5%	1.5%
Basement	**Basement: Yes**	96.5%	97.5%	1.0%
Waterfront	**Waterfront: Yes**	37.5%	76.7%	39.2%
Levels	**Combined**	94.7%	96.0%	1.3%

To see a more detailed breakdown of the offline results click here.

Earlier, we used to receive tags for some listings from the MLS, which were augmented by regex maintained by Data Agg. With the deployment of Smart Tags, we were able to replace the regex approach with the ML models. Now, we use Smart Tags to augment the data we receive from the MLS. Using Smart Tags, we were able to increase our coverage while maintaining high accuracy.

	MLS + Regex (before release - 17 Feb)		MLS + Smart Tags (after release - 19 Mar)	
Tags	**Count**	**Coverage**	**Count**	**Coverage**
Pool	125,518	5.57%	151,860	6.42%
WaterFront	238,154	10.56%	312,772	13.23%
Basement	545,032	24.18%	583,854	24.69%
Garage	1,136,276	50.40%	1,217,677	51.49%
Stories	1,217,653	54.01%	1,429,667	60.45%
Total Active Listings	2,254,316		2,364,848	

Regex Tags

We are generating 106 other tags based purely on regex. Together, these are about 5M tags for 1.9M listings. Top 10 tags for for-sale listings are highlighted below:

Tag	# of listings enhanced (Mar 18)
Big lot	589,904
Floor plan	320,416
Master suite	265,650
Unfinished basement	236,534
Master bedroom	233,076
Farm	205,180
Modern kitchen	202,222
Open floor plan	188,919
Rental property	187,914
Ranch	149,531

What's next

For the next iteration of the 'Smart Tags', fixer-upper homes will be tagged. Also, we are analyzing the impact of different threshold values on the current Smart Tags to improve the precision of the tags (ie. reduce false positives). Another use case of NLP is a free text search that the data science team has started working on the model.

Release Announcement: Machine Learning - Noise Overlay

Realtor.com®'s new noise indicator enables home shoppers to understand a property's noise level before they visit

SANTA CLARA, Calif., March 9, 2020 /PRNewswire/ -- There are many things you can determine from a property listing – a home's size, location, aesthetics, school districts and much more. However, there are some things you simply can't see – such as how noisy or quiet a property may be. Today, realtor.com® introduced a new noise indicator feature that provides reliable sound data down to the property level. Understanding the noise level surrounding a property gives consumers another piece of valuable information they need to make confident buying decisions.

Consumer surveys have found that the most important aspects when considering a new home are price, schools, commute, crime and noise. Realtor.com®'s noise indicator is a first-of-its-kind feature among national real estate search sites providing home shoppers with data at the property level, where others can only do so at a neighborhood level.

"Every home buyer is different. Some people are at home in a bustling city, while others prefer the peace and quiet of a country farmhouse," said Rachel Morley, senior vice president, product management, realtor.com®. "With our new noise indicator, we can provide specific information about whether the property is near a freeway, an airport or a gas station and how that impacts

sound levels. This information can help narrow the search and make sure that our users find a home that's perfect for them."

Noise works in two ways. First, each property will be assigned a noise rating—high, medium or low. Users will also be able to drill down into sources of noise near each property and view details on a heat map overlay that displays noise sources. The feature takes into account three sources of noise: traffic, airports and local sources like restaurants, gas stations, sports stadiums, schools and more. These sources are combined to assign each property a rating. Because the new tool can delineate noise levels down to the individual property level, homes in the same neighborhood may have different ratings based on proximity to major roads, hospitals or schools, for example.

Realtor.com® empowers consumers with extensive property listings and accurate information to help home buyers and sellers make confident and informed decisions.

Click here to try the noise indicator; now available for properties across the continental U.S. on iOS, Android, web and mobile web.

About realtor.com®

Realtor.com® makes buying, selling and living in homes easier and more rewarding for everyone. Realtor.com® pioneered the world of digital real estate 20 years ago, and today through its website and mobile apps is a trusted source for the information, tools and professional expertise that help people move confidently through every step of their home journey. Using proprietary data science and machine learning technology, realtor.com® pairs buyers and sellers with local agents in their market, helping take the guesswork out of buying and selling a home. For professionals, realtor.com® is a trusted provider of consumer connections and branding solutions that help them succeed in today's on-demand world. Realtor.com® is operated by News Corp [Nasdaq: NWS, NWSA] [ASX: NWS, NWSLV] subsidiary Move, Inc. under a perpetual license from the National Association of REALTORS®. For more information, visit realtor.com®.

Media Contact

Nicole Murphy

nicole.murphy@move.com

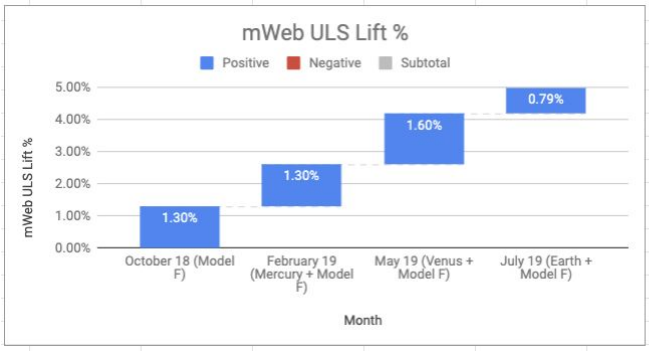
SOURCE realtor.com

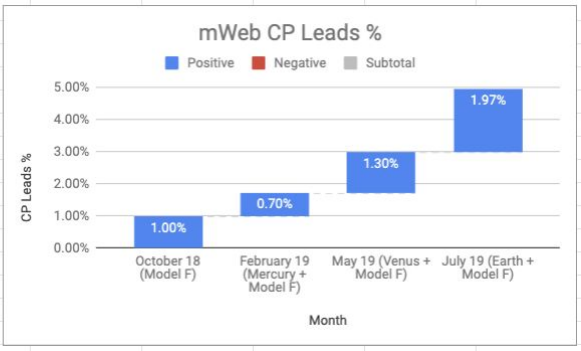
Release Announcement: Machine Learning - Search Ranking on Mobile Apps

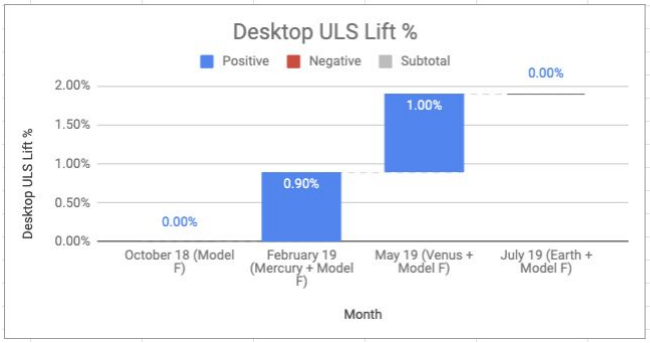
Data Science and Search teams are excited to announce that our newest machine learning models, ModelG and Earth, were released in production on 02/03/2020. Special thanks to everyone who helped and contributed to making this release a great success including Data Science, Search, SRP, A/B test, mAPI, and Homes teams.

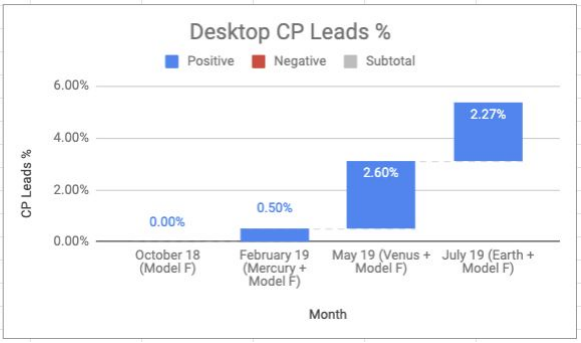
ModelG and Earth are the first models that we have tested on mobile apps. ModelG and Earth increased leads by **~5%** and unique lead submitters by **~2.5%** on iOS; conversion metrics results were inconclusive on Android; however, engagement metrics increased on Android. For example, SRP and LDP saves increased by **~2.2%**.

We released these new search ranking models on Android and iOS mobile apps following the previous four successful releases of our search ranking models on desktop and mWeb which significantly increased leads and unique lead submitters by around **5%** not considering the compounding impact:





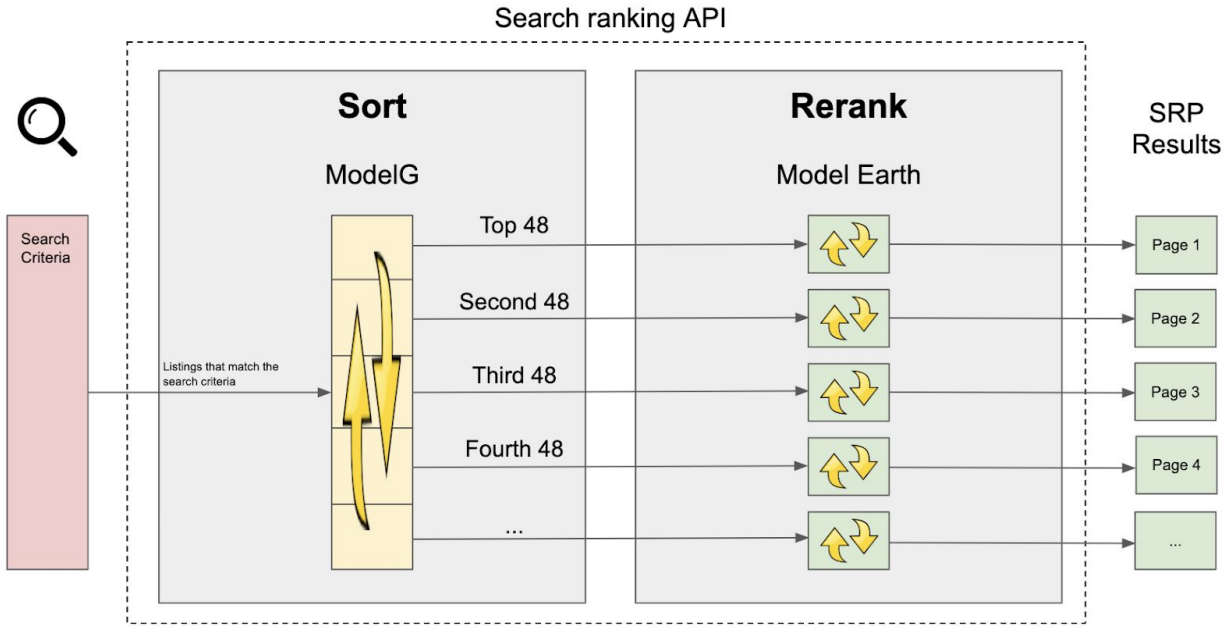




Speed is a primary object of search ranking and led to a two-step approach shown in the diagram below. The first Sort step is implemented using ModelG which provides a fast mechanism for retrieving listings from a pool of listings in Elasticsearch. The retrieved listings are divided into groups of 48 listings to be shown on a SRP page. The second Rerank step is implemented using Model Earth which ranks the listings on each SRP page but does not change their assignment across SRP pages.

ModelG is a MARS (Multivariate Adaptive Regression Splines) model that learns from historically viewed and not viewed listings. ModelG uses listing attributes that are available in ElasticSearch, as well as other engineered features to compare a listing to other listings that received leads in the past, and optimizes for listing views, dwell time and lead submission. Its predecessor, ModelF, is a rule-based model that demotes certain listings based on listing attributes that historically received fewer leads such as freshness, property type, number of photos, etc.

Model Earth is a tree-based algorithm implemented in rankr (ie. rerank), which also learns from historical listing views. Compared with ModelG and its predecessor Model Venus, it uses more features such as location-based features which are highly important in understanding listings' potential value. Incorporating our understanding of location profile turns out to have given us the biggest lift across all iterations.

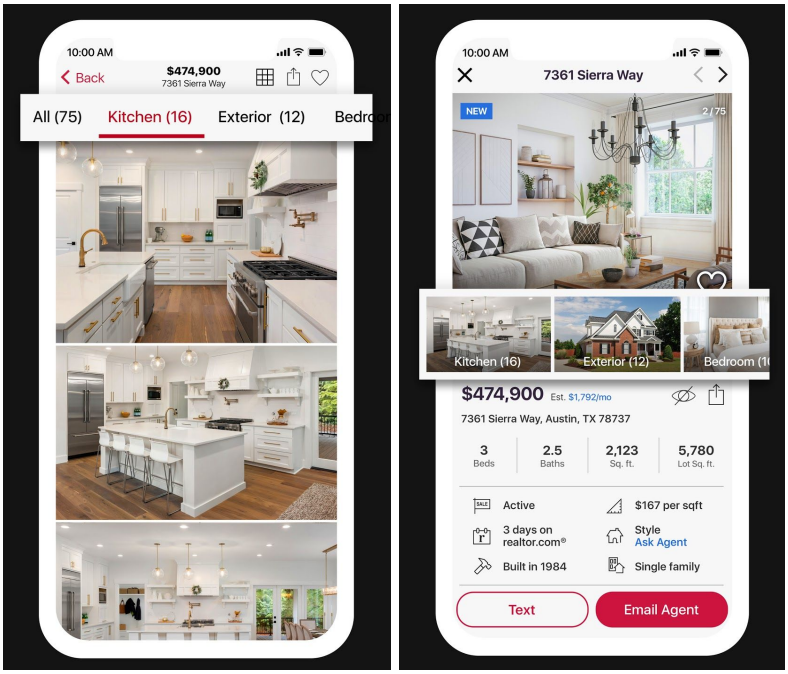


ModelG was optimized by search engineering to take advantage of features available in the Elasticsearch index for quick lookup and execution structure.

Release Announcement: Machine Learning - Photo First

Deep learning technology automatically categorizes and displays photos to simplify home search

SANTA CLARA, Calif., Oct. 29, 2019 /PRNewswire/ -- Realtor.com®, the Home of Home Search℠, today announced the release of its Photo First℠ feature, to help home buyers find their must-have rooms and finishes without the headache of scrolling through dozens of listing photos. The first-of-its-kind feature is designed to deliver the best browsing experience by making photos more personal and relevant than ever before. Powered by deep learning, the Photo First feature can automatically recognize characteristics of a room and organize photos into categories with more than 97 percent accuracy.



Realtor.com®'s Photo First℠ feature automatically categorizes and displays photos to simplify home search.

For consumers, photos are an essential element of a home search. Realtor.com® has reimagined the listing photo experience to make it simpler and more useful than ever before. Now, with the Photo First feature, buyers can quickly hone in on the photos that are most important to them by simply selecting a feature category: exterior, kitchen, bathroom, bedroom,

living room or dining room. And thanks to the feature's deep learning-based algorithm, there is no additional work required for the seller or listing agent to optimize photos.

“Imagine a home search that’s completely customizable. One where you can choose all the features that are important to you, and not only find those homes quickly, but view those photos first,” said Chung Meng Cheong, chief product officer. “The Photo First feature uses deep learning to optimize and categorize listing photos for each home to simplify the home journey. So, if you’re particular about features in your kitchen or bathroom, you can view those photos right away”.

Developed by the mobile and data science teams, the Photo First feature is powered by a proprietary deep learning model for image classification, which enables it to recognize different aspects and automatically associate each photo with the coinciding room. The result is a simple and streamlined interface that users love. In fact, 100 percent of test users preferred the Photo First interface and more than 60 percent engaged with photo categories during testing. The new functionality led to more consumers viewing the home’s details and taking the next step to connect with an agent and learn more.

Realtor.com continues to invest in developing the best-in-class user experience for home shoppers. Today’s announcement is the first of several planned AI-powered photo feature updates designed to help make buying and selling a home simpler and more enjoyable than ever. Photo First is now available on Android and iOS, and will be coming soon to web and mobile web.

Results

In an effort to enhance realtor.com’s consumer experience using a photo-centric approach, data science team developed an image tagging solution that tags listing images into 16 different categories. This new Image Tagger has been developed using a deep learning approach to achieve an accuracy of 97.4%, a big improvement from 92% accuracy of the previous model, which tagged images in only 8 categories.

The first use case of this new image tagger was the Photo First project; which has led to significant increase in our engagement metrics.

- LDP
    - **+16%** Android LDP Page Views / Unique User
    - **+18%** Android SRP to LDP Conversions / Unique User
- Photo Gallery
    - **+8%** lift in unique users entering the photo gallery
    - **+22%** lift in taps per unique user to enter the photo gallery
- Photo Gallery Monetization
    - **+22.5%** Android leads from photo gallery in-line lead form
    - **+12.6%** Android leads from photo gallery bottom email lead button